

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO ENGENHARIA GEOGRÁFICA, GEOFÍSICA E ENERGIA



**Utilização de dados geográficos voluntários e
técnicas de deteção remota na atualização de redes
viárias em áreas protegidas**

David Miguel de Sousa Francisco

Mestrado em Sistemas de Informação Geográfica – Tecnologias e Aplicações

Trabalho de Projeto orientado por:
Prof. Doutor João Catalão Fernandes
Dr. Ricardo Nogueira Mendes

“Estamos a entrar num novo mundo, no qual os dados podem ser mais importantes que o software”

Tim O'Reilly

AGRADECIMENTOS

No término de mais uma etapa da minha vida académica, não posso deixar de dar um agradecimento muito especial aos meus orientadores, Professor Doutor João Catalão Fernandes e ao Dr. Ricardo Nogueira Mendes, muito obrigado por todo o apoio e orientação que me forneceram.

Deixar um agradecimento à minha família e em especial à minha mãe que sempre me apoiou e sem ela, certamente não tinha chegado a este ponto.

A todos os meus sinceros agradecimentos por toda ajuda e apoio.

Muito Obrigado!

RESUMO

Este trabalho de projeto incide sobre a área geográfica do Parque Natural da Arrábida (PNArr), uma zona muito frequentada pela população, especialmente nas atividades de ciclismo/BTT e passeios pedestres. A grande afluência ao parque a juntar à falta de sinalização, levou a uma proliferação dos trilhos existentes, destruindo desta forma uma flora rica e única.

O principal objetivo deste trabalho de projeto consiste na criação de uma rede viária que contemple todos os trilhos existentes, bem como a rede viária já existente e cartografada. Para se atingir o objetivo proposto, foram utilizados dados geográficos voluntários, técnicas de deteção remota e uma imagem do satélite Sentinel-2, nas bandas do azul, verde, vermelho e infravermelho próximo, todas com uma resolução espacial de 10m.

Os dados geográficos voluntários representam os trilhos efetuados por cada utilizador. Foi efetuada uma máscara com esses mesmos trilhos, com o intuito de auxiliar na classificação da imagem de satélite, recorrendo a quatro algoritmos distintos: Máxima verossimilhança; SVM; Rede Neuronal; Árvore de decisão. Para cada algoritmo com exceção da árvore de decisão foram efetuadas duas classificações, uma sem o auxílio da máscara com os trilhos, outra com o auxílio desses dados. Após as classificações, foram realizadas as matrizes de confusão, de modo, a analisar os resultados obtidos, com o âmbito de verificar se realmente existia uma melhoria na classificação utilizando a máscara com os trilhos provenientes dos dados geográficos voluntários.

Os resultados obtidos são satisfatórios, mas existe margem para melhorar, tanto na classificação das imagens, como na resolução espacial da imagem de satélite.

PALAVRAS CHAVE: Dados Geográficos Voluntários, Deteção Remota, Sentinel-2, Algoritmos de classificação de imagem, Parque Natural da Arrábida

ABSTRACT

This project work focuses on the geographical area of Arrabida Natural Park, an area with a huge demand specially for mountain bike and hiking lovers. The large affluence to the park coupled with the lack of signage led to a proliferation of existing trails, thus destroying a unique and rich flora.

The main goal of this project work is to create a road network that covers all the existing trails as well as the existing and mapped network.

To achieve the proposed objective, voluntary geographic information, remote sensing techniques and a Sentinel-2 satellite imagery were used, four bands were used: Blue, green, red and near infrared, all with a 10m special resolution.

The voluntary geographic information represents the trails made by each user. A mask was created with these same trails, in order to assist the classification of the satellite imagery, using four distinct algorithms: Maximum Likelihood; SVM; Neural Network; Decision Tree. For each algorithm except the decision tree, two classifications were made, one without aid of the mask with the trails, the other one with the aid of the mask.

After the classifications, confusions matrices were performed in order to analyze the results obtained and if there really was an improvement in the classification where the mask with the trails were used.

The results were reasonable, but like always there is room to improve, both on image classification and on the spatial resolution of the imagery.

KEYWORDS: Voluntary Geographic Information, Remote Sensing, Sentinel-2, Image Classification Algorithms, Arrabida Natural Park

ÍNDICE

AGRADECIMENTOS	iii
RESUMO.....	iv
ABSTRACT.....	v
ÍNDICE.....	vi
LISTA DE TABELAS	viii
LISTA DE FIGURAS.....	ix
ACRÓNIMOS.....	x
CAPÍTULO 1	1
INTRODUÇÃO	1
1.1. ENQUADRAMENTO	1
1.2. OBJETIVOS.....	2
1.3. ORGANIZAÇÃO DO TRABALHO	3
CAPÍTULO 2	4
ESTADO DA ARTE.....	4
CAPÍTULO 3	7
DADOS E MÉTODOS	7
3.1. ÁREA DE ESTUDO	7
3.2. DADOS UTILIZADOS	8
3.2.1. Imagem de Satélite Sentinel-2	8
3.2.2. Dados Voluntários	9
3.3 MÉTODOS.....	11
3.3.1. Recolha e Tratamento da imagem de satélite Sentinel-2.....	12
3.3.2. Classificação de Imagem	14
a) Máxima Verosimilhança.....	14
b) <i>Support Vector Machine</i>	14
c) Redes Neurais	15
d) Árvore de Decisão	15
3.3.3. Validação dos Resultados Obtidos	17
CAPÍTULO 4	18
RESULTADOS E DISCUSSÃO	18

4.1 TRATAMENTO DOS DADOS GEOGRÁFICOS VOLUNTÁRIOS	18
4.2 CLASSIFICAÇÃO DE IMAGEM.....	20
4.3 RESULTADOS DO ALGORITMO DE MÁXIMA VEROSIMILHANÇA....	23
4.4. RESULTADOS DO ALGORITMO SVM.....	25
4.5. RESULTADOS DO ALGORITMO REDE NEURONAL.....	27
4.6. RESULTADOS DO ALGORITMO ÁRVORE DE DECISÃO	30
4.7. COMPARAÇÃO ENTRE OS ALGORITMOS.....	32
CAPÍTULO 5	34
CONCLUSÃO E PERSPETIVAS FUTURAS.....	34
REFERÊNCIAS BIBLIOGRÁFICAS.....	35
ANEXOS.....	37
A. Imagens classificadas	37
i. Classificação através do algoritmo máxima verossimilhança sem dados auxiliares	37
ii. Classificação através do algoritmo máxima verossimilhança com dados auxiliares	37
iii. Classificação através do algoritmo SVM sem dados auxiliares.....	38
iv. Classificação através do algoritmo SVM com dados auxiliares	38
v. Classificação através do algoritmo rede neuronal sem dados auxiliares	39
vi. Classificação através do algoritmo rede neuronal com dados auxiliares	39
vii. Classificação através do algoritmo árvore de decisão com dados auxiliares	40
B. Matrizes de Confusão Completas.....	41
i. Matriz de confusão máxima verossimilhança sem dados auxiliares	41
ii. Matriz de confusão máxima verossimilhança com dados auxiliares.....	42
iii. Matriz de confusão SVM sem dados auxiliares	43
iv. Matriz de confusão SVM com dados auxiliares.....	44
v. Matriz de confusão rede neuronal sem dados auxiliares	45
vi. Matriz de confusão rede neuronal com dados auxiliares	46
vii. Matriz de confusão árvore de decisão com dados auxiliares	47

LISTA DE TABELAS

Tabela 3.1. Bandas espectrais Sentinel-2 (ESA)	9
Tabela 3.2. Softwares utilizados e respectivos processos	10
Tabela 4.1 Percentagem dos dados eliminada.	18
Tabela 4.2. Número de pixéis por classe nas áreas de treino e validação	21
Tabela 4.3. Número de cada banda no Layer Stacking	21
Tabela 4.4. Matriz de confusão para a classificação com o algoritmo de Máxima Verosimilhança sem dados auxiliares	24
Tabela 4.5. Matriz de confusão para a classificação com o algoritmo de Máxima Verosimilhança com dados auxiliares	25
Tabela 4.6. Matriz de confusão para a classificação com o algoritmo SVM sem dados auxiliares.....	26
Tabela 4.7. Matriz de confusão para a classificação com o algoritmo SVM com dados auxiliares.....	27
Tabela 4.8. Matriz de confusão para a classificação com o algoritmo Rede Neuronal sem dados auxiliares	29
Tabela 4.9. Matriz de confusão para a classificação com o algoritmo Rede Neuronal sem dados auxiliares	29
Tabela 4.10 Matriz de confusão para a classificação com o algoritmo árvore de decisão, com dados auxiliares	31

LISTA DE FIGURAS

Figura 1.1. Fotografia de alguns trilhos no PNArr	2
Figura 3.1. Parque Natural da Arrábida ICNF	7
Figura 3.2 Dados brutos (vetoriais)	10
Figura 3.3. Diagrama dos processos efetuados	12
Figura 3.4. Imagem Sentinel-2 da área de estudo	13
Figura 3.5. Diagrama da primeira fase do projeto	13
Figura 3.6. Diagrama da segunda fase do projeto	14
Figura 3.7. Demonstração do algoritmo SVM	15
Figura 3.8 Diagrama de funcionamento de um nó da rede neuronal.....	15
Figura 3.9. Classificação através do algoritmo Árvore de Decisão (www.psckhub.com)	16
Figura 3.10 Diagrama da terceira fase do projeto	16
Figura 4.1. Máscara dos trilhos	19
Figura 4.2 Máscara dos trilhos ampliada (Cabo Espichel/Sesimbra)	19
Figura 4.3. Modelo Digital de Terreno.....	20
Figura 4.4. Gráfico com a assinatura espectral por classe em cada banda	22
Figura 4.5. Árvore de Decisão com os critérios utilizados para a classificação da imagem	22
Figura 4.6. Classificação Máxima Verossimilhança sem dados auxiliares	23
Figura 4.7. Classificação Máxima Verossimilhança com dados auxiliares.....	24
Figura 4.8. Classificação SVM sem dados auxiliares	25
Figura 4.9. Classificação SVM com dados auxiliares	26
Figura 4.10. Classificação com algoritmo Rede Neuronal sem dados auxiliares.....	28
Figura 4.11. Classificação com algoritmo Rede Neuronal com dados auxiliares	28
Figura 4.12. Classificação através do algoritmo de Árvore de Decisão com dados auxiliares ...	30
Figura 4.13. Classificação dos trilhos e das vias artificiais	31

ACRÓNIMOS

BTT	Bicicleta Todo o Terreno
DR	Deteção Remota
ESA	<i>European Space Agency</i>
GNSS	<i>Global Navigation Satellite System</i>
IGV	Informação Geográfica Voluntária
MDT	Modelo Digital de Terreno
NDVI	<i>Normalized Difference Vegetation Index</i>
OSI	<i>Ordinance Survey Ireland</i>
OSM	<i>Open Street Map</i>
PNArr	Parque Natural da Arrábida
S2A	Sentinel-2 A
SIG	Sistemas de Informação Geográfica
SVM	<i>Support Vector Machine</i>
VGI	<i>Voluntary Geographic Information</i>

CAPÍTULO 1

INTRODUÇÃO

1.1. ENQUADRAMENTO

A pressão de utilização recreativa do Parque Natural da Arrábida nos dias de hoje tem crescido a um ritmo acelerado, refletindo-se na proliferação de caminhos informais e ilegais com vários impactos, que colocam em risco a biodiversidade e o ecossistema do Parque, não existindo uma atualização destas infraestruturas que permita a sua melhor gestão no âmbito dos objetivos dos planos de ordenamento das áreas protegidas.

O Parque Natural da Arrábida (PNArr) foi criado em 1976 com o propósito de proteger os valores geológicos, florísticos, faunísticos e paisagísticos locais (ICNF, 2019). O PNArr estende-se da cidade de Setúbal a este até ao Cabo Espichel a oeste. Em 1998, a delimitação do PNArr foi ampliando com a criação da área marinha Arrábida-Espichel, completando no meio marinho os objetivos de conservação da natureza subjacentes ao Parque (ICNF, 2019).

A proximidade da área metropolitana de Lisboa a par da crescente procura de zonas recreativas naturais coloca o PNArr sob grande pressão pelos amantes da natureza criando e recriando trilhos e percursos com evidente interferência no ecossistema. Na atualidade, a sinalização existente é muito reduzida ou nenhuma, o que leva a que os praticantes que lá se deslocam não tenham noção de quais as rotas a seguir, consequentemente leva ao aumento de trilhos ilegais e à consequente destruição da flora. É importante recordar que existem áreas no PNArr que só é possível serem acedidas quando acompanhadas por um guia indicado pela sede do Parque.

As atividades mais praticadas no PNArr são o ciclismo, BTT e pedestrianismo. Muitos dos praticantes destas atividades recreativas registam voluntariamente os seus percursos com equipamentos GNSS e disponibilizam em sítios da web dedicados. Estes dados, dados geográficos voluntários, são uma fonte de informação relevante como indicadores da intensidade de utilização e no mapeamento das atividades recreativas. A ideia que nos propomos explorar neste projeto é a da utilização dos dados geográficos voluntários na atualização da rede viária dentro das áreas protegidas, usando dados de satélite e técnicas de deteção remota para melhorar a qualidade de informação sobre as mesmas e posteriormente poder gerir e monitorizar as redes viárias.

1.2. OBJETIVOS

O objetivo deste trabalho projeto é aliar o acesso gratuito e fácil aos dados geográficos voluntários e as técnicas de deteção remota, com o intuito de criar uma rede viária atualizada em áreas protegidas, que no caso específico deste trabalho projeto incide sobre o Parque Natural da Arrábida.

No sentido de construir a rede viária do Parque Natural da Arrábida, os dados geográficos voluntários das atividades recreativas mais praticadas são usados para criar uma camada de informação com dados relativos à localização dos trilhos e à sua intensidade de utilização. Esta camada de informação será combinada com informação espectral do satélite Sentinel-2 e com informação da morfologia do terreno (altimetria, declive e orientação) com o objetivo de classificar as vias em tipologias de piso. Para o efeito serão testados algoritmos de classificação automática, como o algoritmo de Máxima Verosimilhança, *Support Vector Machine* (SVM), Rede Neuronal e Árvore de Decisão. O resultado do projeto será a rede viária atual do parque, bem como uma proposta metodológica para a construção do mesmo produto em territórios de contexto semelhante.



Figura 1.1. Fotografia de alguns trilhos no PNArr

1.3. ORGANIZAÇÃO DO TRABALHO

Este trabalho projeto está dividido em 5 capítulos, o primeiro subdividido em três pontos. O segundo capítulo contém o estado de arte, onde são apresentadas algumas técnicas utilizadas por outros autores em projetos semelhantes.

O capítulo 3 integra alguns conceitos teóricos fundamentais na elaboração e percepção deste trabalho projeto.

No capítulo 4 é são apresentados os resultados obtidos, através da classificação por cada algoritmo utilizado, e uma breve discussão sobre cada resultado.

O capítulo 5 é dedicado às conclusões deste trabalho projeto, dificuldades encontradas e as perspectivas futuras.

CAPÍTULO 2

ESTADO DA ARTE

Na última década assistimos a um aumento generalizado da aquisição e divulgação de dados geográficos voluntários com informação sobre diversas temáticas como as estradas, trilhos, ocupação do solo, que têm alimentado alguns sítios de mapas colaborativos e serviços baseados na localização. Um dos mais conhecidos repositórios de informação geográfica voluntária é o *Open Street Map* (OSM).

O projeto *Open Street Map*, tem como objetivo criar um mapa digital grátis que abranja todas as zonas do mundo. Recorre a dados geográficos voluntários, qualquer pessoa pode contribuir para o projeto de uma forma voluntária. O principal problema dos dados geográficos voluntários é a ausência de um mecanismo de autoavaliação ou de supervisão da qualidade posicional e temática dos dados. Várias entidades ligadas ao mundo dos SIG, reafirmam a necessidade de averiguar a credibilidade e qualidade da informação. Até certa altura vários profissionais ligados à área dos SIG, estavam bastante relutantes em aceitar como credível informação proveniente de dados geográficos voluntários, tendo duas premissas, a falta de medidas de controlo na sua fiabilidade e precisão e visto se tratar de uma plataforma de dados voluntários, a maior parte dos seus contribuintes serem “amadores”.

Uma das técnicas utilizadas para determinar a fiabilidade e credibilidade da informação é a comparação das formas dos polígonos entre o OSM e o *Ordnance Survey Ireland* (OSI), para isso implementaram a *turning-function shape similarity metric* (Arkin et al., 1990). Neste estudo, os autores concluíram que o OSM é um projeto muito atrativo para a área dos SIG, em grandes cidades é funcional e até mesmo uma opção bastante viável, no entanto o mesmo não acontece em áreas rurais, ou mais pequenas, onde a quantidade de informação não pode ser comparada à existente em cidades mais conhecidas e maiores, tendo mesmo obtido baixos níveis de credibilidades nessas áreas (Mooney et al.).

Até que ponto será possível confiar na fiabilidade dos dados geográficos voluntários? Existe uma grande heterogeneidade de dados, comparativamente aos dados das grandes agências nacionais de mapeamento. Em Londres, Inglaterra, o OSM existe desde agosto de 2004, foi elaborada uma comparação entre os dados do OSM e das bases de dados do próprio país. OSM obteve uma precisão de cerca de 6m e uma sobreposição na ordem dos 80%, no que diz respeito a vias principais na cidade de Londres. No espaço temporal de quatro anos o OSM, mapeou cerca de 29% do território Inglês (Haklay 2008). O OSM apresenta uma correspondência que varia entre o 100% e os 50%,

comparativamente com os dados recolhidos pelas agências oficiais de mapeamento dos países onde foi efetuado um estudo sobre a qualidade e precisão dos dados geográficos voluntários, podendo mesmo apresentar valores inferiores aos 50% em alguns casos. Isto deve-se em grande parte à enorme variedade de utilizadores, que na sua maioria são amadores. Na teoria, quanto maior for o número de utilizadores, melhor será o sistema, é provável que com um maior número de utilizadores, sejam colmatadas algumas incorreções que possam existir (Haklay *et al.*, 2010).

Foi elaborado um estudo com o intuito de avaliar e comparar os dados do OSM, na Grécia. Os dados geográficos voluntários presentes no OSM foram comparados, com os dados do Serviço Geográfico Militar Grego (HMGS – *Hellenic Military Geographical Service*), o órgão oficial de cartografia do país da Grécia. Os resultados no geral foram bastante satisfatórios, no que diz respeito ao comprimento dos dados, houve uma correspondência de 88%, a pesquisa pelo nome teve uma correspondência de 87%, e a percentagem média de sobreposição foi de 89%. O nome completo e o tipo obtiveram uma correspondência bastante inferior, às registadas anteriormente, 26% e 33%, respetivamente (Kounadi 2009).

O aumento da procura na utilização de dados geográficos voluntários, levou a que fosse necessário criar várias ferramentas, com o intuito de averiguar a qualidade desses mesmos dados. Atualmente vários programas de sistemas de informação geográfica já vêm dotados de algumas destas ferramentas. Isto demonstra que os dados geográficos voluntários estão de facto a avançar dos conceitos académicos e estão a ser postos em prática e a ser utilizados atualmente, com diversas finalidades (Hunter 1999).

A evolução da internet, levou a uma explosão de interesses, várias organizações estão a encorajar os seus utilizadores a fornecerem dados e aplicações que considerem do seu interesse voluntariamente, como é o caso do *Open Street Map* e da *Wikimapia*, que estão a incentivar os seus utilizadores a construir uma rede geográfica global. O *Google Earth* e outras organizações semelhantes, estão a incentivar os seus utilizadores a desenvolverem aplicações interessantes utilizando os seus próprios dados (Goodchild, 2007). O grande motivo pelo qual existe uma controvérsia na utilização e no encorajamento nos dados geográficos voluntários, é que pode ser uma atividade considerada altamente exploratória. Os utilizadores são encorajados a participar num grande projeto em prol de um bem maior, o mapeamento global, sem receber absolutamente nada por isso, na verdade vários autores alegam que na verdade, existe uma organização que beneficia e lucra com esses dados. Em certos casos, o mais barato para as empresas é a utilização de dados voluntários, o que significa de imediato um corte nas despesas dessa empresa, no qual o trabalho que cada utilizador tem é utilizado para diversos projetos e trabalhos sem qualquer recompensa ou obrigação entre a empresa e o “empregado” (Haklay 2008).

Os dados geográficos voluntários oferecem uma fonte barata de dados georreferenciados, capazes de ser utilizados na monitorização da cobertura do solo

utilizando técnicas de detecção remota (Stehman et al. 2018). A qualidade das observações e a falta de um projeto de amostragem probabilística, levantam preocupações sobre o uso de informação geográfica voluntária. Existem algumas opções para tornar os dados mais fidedignos, instruir os voluntários a obter dados para locais selecionados, aumentar os dados a partir de uma amostra probabilística, utilizar os dados geográficos voluntários para criar uma variável auxiliar que irá ser usada como estimador assistido por um modelo, com o objetivo de reduzir o erro padrão de uma estimativa produzida através de uma amostra probabilística.

A produção de mapas da cobertura do solo, são produtos fundamentais, com uma vasta gama de aplicações, a produção destes mapas ao longo do tempo, pode também ser usada para detetar as alterações do solo (Brovelli et al. 2018). O processo baseia-se na classificação das imagens de satélite e a validação dos dados, é fundamental verificar a precisão dos dados e da classificação antes de serem utilizados em aplicações reais, como a proteção e monitorização da natureza e biodiversidade, gestão de recursos naturais, modelos hidrográficos, distribuição de espécies e avaliações ambientais (Foley et al. 2005) e (Nie et al. n.d.).

CAPÍTULO 3

DADOS E MÉTODOS

Neste capítulo irá ser feita uma descrição da área de estudo, bem como dos dados e programas utilizados, a metodologia e processos que foram implementados do projeto.

3.1. ÁREA DE ESTUDO

A área em estudo no qual este trabalho projeto foi aplicado, é o Parque Natural da Arrábida, uma área protegida, devido à sua única flora, à sua extensa fauna, história e cultura. O Parque Natural da Arrábida ocupa cerca de 17mil ha, dos quais 5 mil são de superfície marinha, distribui-se geograficamente por três concelhos, Palmela, Sesimbra e Setúbal, todos pertencentes à Península de Setúbal, que por sua vez integra a área metropolitana de Lisboa.

O maciço Arrábico, desde cedo levou a que se tivessem iniciado algumas tentativas de proteção, devido às características particulares que este apresenta. A 28 de julho de 1976, com a publicação do Decreto-Lei nº622/76, que “reconhecendo a insuficiente proteção conferida pelas medidas preventivas decretadas para a zona...” é criado o Parque Natural da Arrábida (PNArr) (“Decreto Lei nº622/76” 1976). Esta classificação teve como objetivo proteger os valores geológicos, florísticos, faunísticos e paisagísticos locais, bem como testemunhos materiais de ordem cultural e histórica.



Figura 3.1. Parque Natural da Arrábida ICNF

A cordilheira da Arrábida é constituída por três eixos:

- O 1º composto por pequenas elevações nos arredores de Sesimbra, pelas serras do Risco e da Arrábida e pelas colinas existentes entre Outão e Setúbal;
- O 2º é formado pelas Serras de S. Luís e dos Gaiteiros;
- O 3º formado pelas Serras do Louro e de São Francisco.

3.2. DADOS UTILIZADOS

No âmbito da realização deste trabalho de projeto, foram utilizados vários dados distintos, que combinados com várias técnicas de SIG e de deteção remota, pretendem responder aos objetivos propostos.

3.2.1. Imagem de Satélite Sentinel-2

O Sentinel-2 é um projeto europeu (ESA – *European Space Agency*), constituído por dois satélites, com uma largura de faixa ampla (290 km), alta resolução e multiespectrais. Transporta um instrumento ótico, capaz de fornecer treze bandas espectrais (Tabela 3.1) distintas, quatro bandas com uma resolução espacial de 10m, seis bandas com uma resolução espacial de 20m e três bandas com uma resolução espacial de 60m. Os dois satélites têm uma órbita polar e estão colocados na mesma órbita síncrona ao sol desfasados de 180°. Os dois satélites contribuem com as suas bandas multiespectrais para variados serviços e aplicações como, gestão de uso do solo, agricultura e silvicultura, controlo de desastres naturais, intervenções de resgate humanitário, mapas de risco e segurança. Esta missão tem uma resolução temporal de 10 dias no equador com um satélite e de cinco dias com os dois satélites, nas latitudes médias a resolução temporal é de dois a três dias. Os limites da cobertura variam entre as latitudes 56° sul e os 84° norte.

Com 13 bandas espectrais, 290km de largura da faixa de cobertura, e um período de revisita alto, o satélite pode ser incluído numa variedade de estudos e programas, e reduzir o tempo para construir um arquivo de imagens da Europa sem nuvens. As suas bandas espectrais vão fornecer imagens para a deteção e alteração da classificação do uso do solo, correções atmosféricas e a distinção entre nuvens e neve. Esta missão tem como objetivos principais fornecer:

- Imagens sistemáticas multiespectrais de alta resolução aliadas e uma resolução temporal de dois a três dias nas regiões de latitude média;
- Continuidade de imagens multiespectrais fornecidas pela serie de satélites SPOT e pelo instrumento USGS LANDSAT *Thematic Mapper*.
- Recolher dados para a próxima geração de produtos operacionais, como mapas de uso do solo, mapas de alterações do solo e variáveis geofísicas.

Estes são os objetivos prioritários no que diz respeito ao Sentinel-2, o que leva a uma contribuição significativa para o projeto Copernicus nos seguintes temas: alterações climáticas, monitorização do solo, gestão de segurança e emergência (Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Bargellini 2012)

Tabela 3.1. Bandas espectrais Sentinel-2 (ESA)

Bandas Sentinel-2	Comprimento de onda central (μm)	Resolução (m)
Banda 1 - Aerossol	0,443	60
Banda 2 - Azul	0,49	10
Banda 3 - Verde	0,56	10
Banda 4 - Vermelho	0,665	10
Banda 5 - Red edge 1	0,705	20
Banda 6 - Red edge 2	0,74	20
Banda 7 - Red edge 3	0,783	20
Banda 8 - Infravermelho próximo	0,842	10
Banda 8A - Red edge 4	0,865	20
Banda 9 - Vapor de água	0,945	60
Banda 10 - SWIR-Cirrus	1,375	60
Banda 11 - SWIR 1	1,61	20
Banda 12 - SWIR 2	2,19	20

Na elaboração deste trabalho projeto foram utilizadas quatro bandas do Sentinel-2, todas com uma resolução espacial de 10m. Foram utilizadas as seguintes bandas:

- Banda 2 – Azul
- Banda 3 – Verde
- Banda 4 – Vermelho
- Banda 8 – Infravermelho próximo

3.2.2. Dados Voluntários

Foram utilizados dados geográficos voluntários provenientes de duas fontes distintas, a rede viária foi extraída do OSM, enquanto que os trilhos foram descarregados do GPSies (www.gpsies.com). A utilização destes dados, principalmente os trilhos, demonstra quais os caminhos utilizados pelos caminhantes e ciclistas, que praticam as suas atividades de lazer no Parque Natural da Arrábida. É possível identificar os trilhos “clandestinos” que são formados regularmente por alguns dos utilizadores do parque.

Os dados recolhidos, fornecem informação desde o ano 2006 ao ano 2017. Foram utilizados cerca de 50% dos dados, após a validação e tratamento dos dados em bruto, o que corresponde a 1644 trilhos.

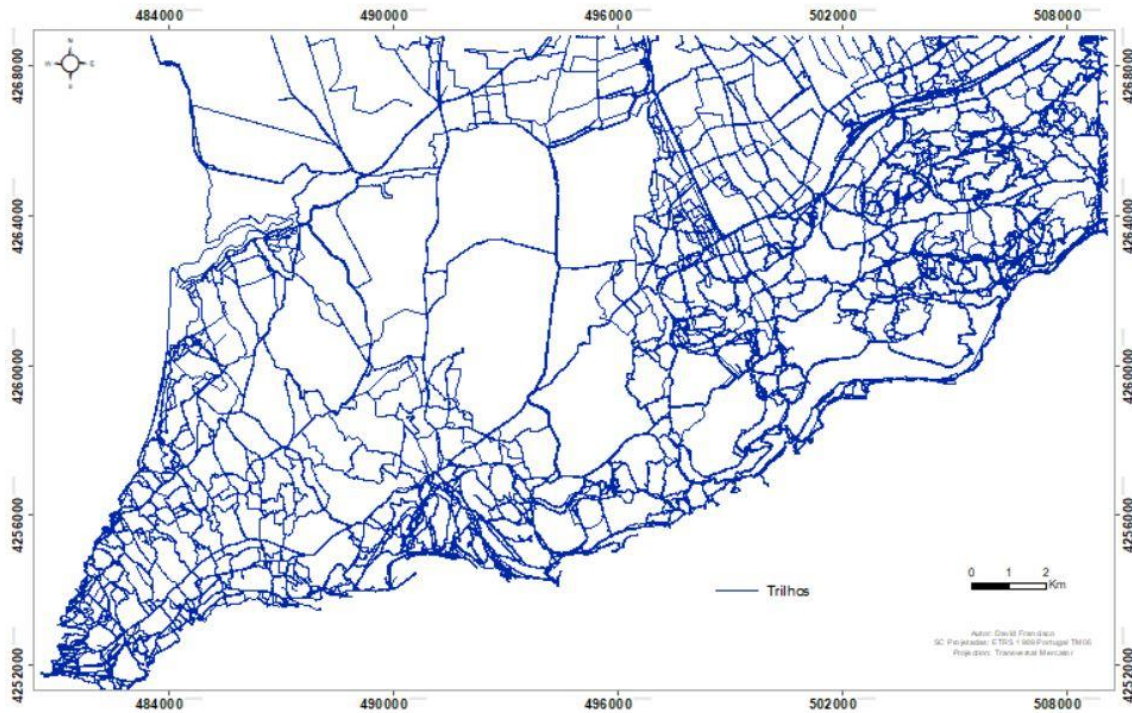


Figura 3.2 Dados brutos (vetoriais)

O processamento dos dados recolhidos exigiu a utilização de diversas aplicações informáticas, nomeadamente ESA SNAP®, ArcGIS 10.4®, ENVI 5.3®, MATLAB R2015a®, e Microsoft® e Excel 365 (Tabela 3.2).

Tabela 3.2. Software utilizado e respetivos processos

SOFTWARE/PROGRAMA	PROCESSOS
ESA SNAP®	<ul style="list-style-type: none"> • Composição colorida da imagem Sentinel-2. • Cálculo do NDVI.
ENVI 5.3®	<ul style="list-style-type: none"> • <i>Layer Stacking</i> de todas as bandas utilizadas. • Utilização dos algoritmos para classificar a imagem.
ARCGIS 10.4®	<ul style="list-style-type: none"> • Tratamento dos dados vetoriais. • Criação de uma máscara com os

	<p>trilhos de bicicleta e com os trilhos a pé.</p> <ul style="list-style-type: none"> • Utilização do modelo digital de terreno, para criar um mapa de declives e exposição solar da área em estudo.
PYTHON 2.7	<ul style="list-style-type: none"> • Programa utilizando a linguagem <i>python</i> para a criação da máscara de trilhos de bicicleta e a pé no ArcGis.
MICROSOFT EXCEL	<ul style="list-style-type: none"> • Tratamento dos dados finais.

3.3 MÉTODOS

Este trabalho de projeto alia duas técnicas distintas, de forma a que as mesmas se complementem, com o objetivo de chegar a um resultado final satisfatório no âmbito do que são os objetivos propostos. A realização de todos os processos (Figura 3.3) pode ser dividida em quatro fases:

1. Recolha e tratamento da imagem de satélite Sentinel-2;
2. Recolha e tratamento dos dados vetoriais e do modelo digital de terreno;
3. Junção das máscaras com os trilhos e classificação das imagens utilizando quatro algoritmos diferentes;
4. Validação dos resultados obtidos.

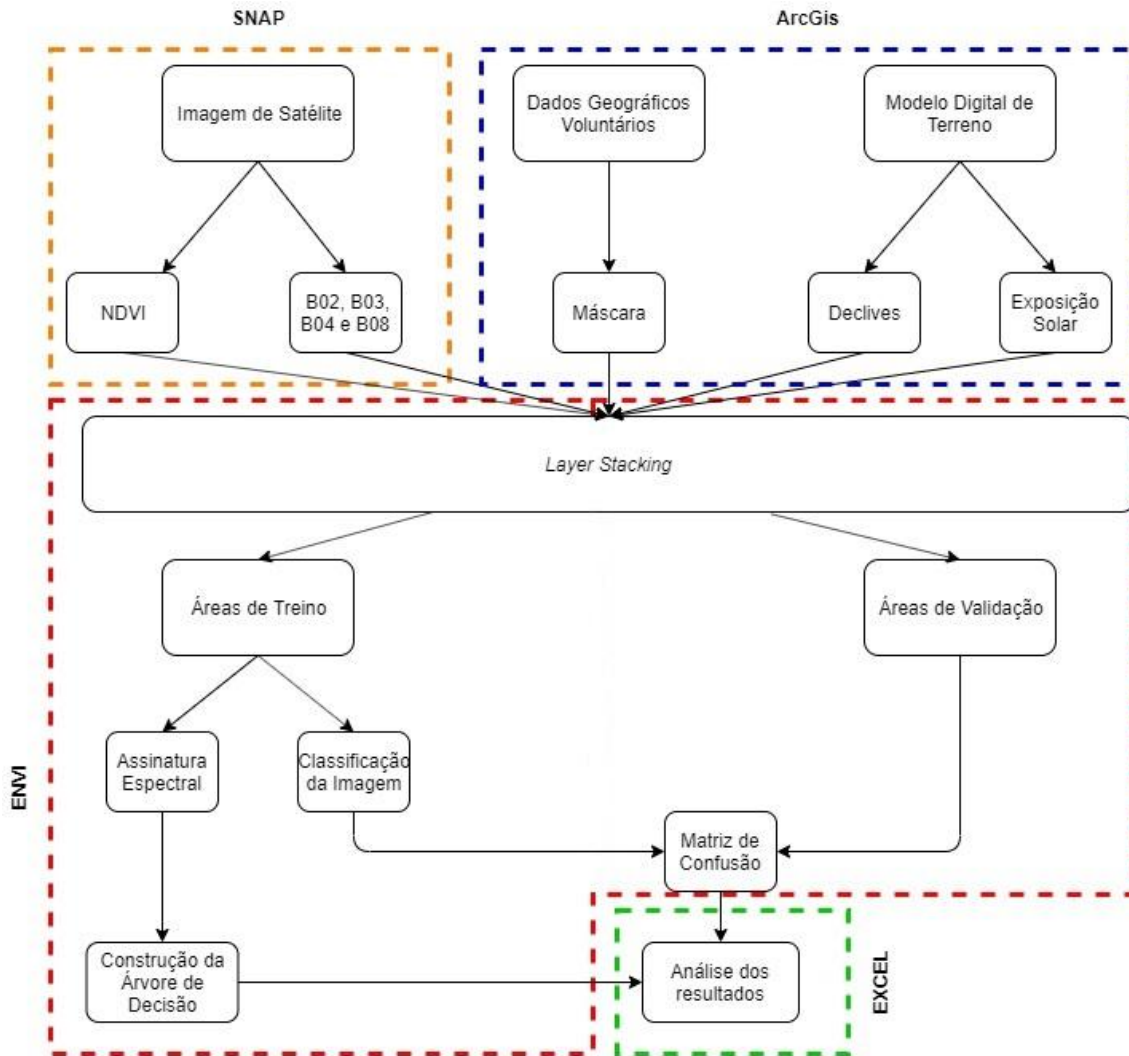


Figura 3.3. Diagrama dos processos efetuados

3.3.1. Recolha e Tratamento da imagem de satélite Sentinel-2

Nesta primeira fase foi feita uma pesquisa pelo repositório de imagens do projeto Copernicus, onde foi selecionada e extraída uma imagem do satélite Sentinel-2, datada de cinco de abril de 2017.

Foram utilizadas quatro bandas, B02, B03, B04 e B08, correspondentes ao azul, verde, vermelho e infravermelho próximo respetivamente. Foi calculado o índice de vegetação NDVI (Normalized Difference Vegetation Index), que utiliza a razão da diferença normalizada entre as bandas do infravermelho próximo e do vermelho (Rouse et al. 1973), como mostra a equação (1).

$$NDVI = \frac{\rho_{NIR} - \rho_{red}}{\rho_{NIR} + \rho_{red}} \quad (1)$$

Este índice tem como objetivo avaliar a atividade fotossintética, os seus valores variam entre -1 e 1, e representam a refletância, valores iguais ou menores que zero indicam ausência de vegetação, sendo que os valores negativos representam água e os valores próximos de zero representam solo nu. Quanto mais perto do valor um, maior é a quantidade de vegetação fotossinteticamente ativa.

Após o cálculo do índice de vegetação NDVI, foi criada uma composição colorida cor natural (Figura 3.4), utilizando as três bandas correspondentes a região do visível no espectro eletromagnético, B02, B03 e B04. Foi também efetuado um corte na imagem de satélite original, de forma a ser representada a área de estudo.

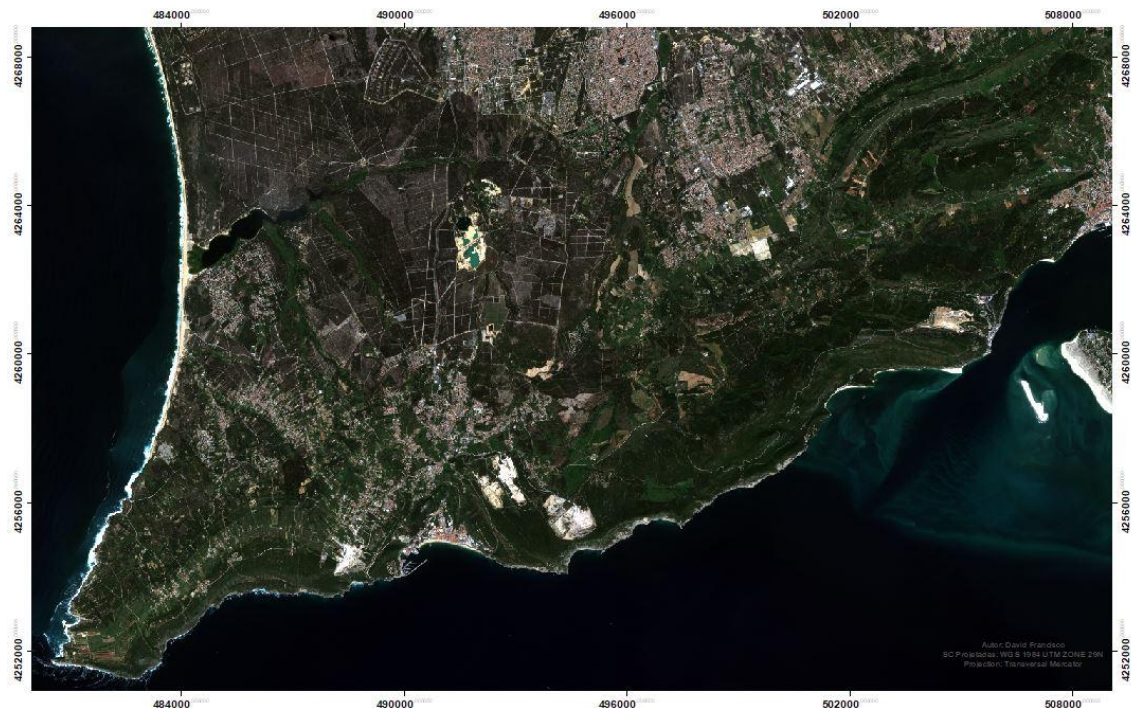


Figura 3.4. Imagem Sentinel-2 da área de estudo

A primeira fase do projeto (Figura 3.5) baseou-se na recolha da imagem de satélite Sentinel-2, e posteriormente foi efetuado um corte sobre a área incidente deste trabalho projeto.

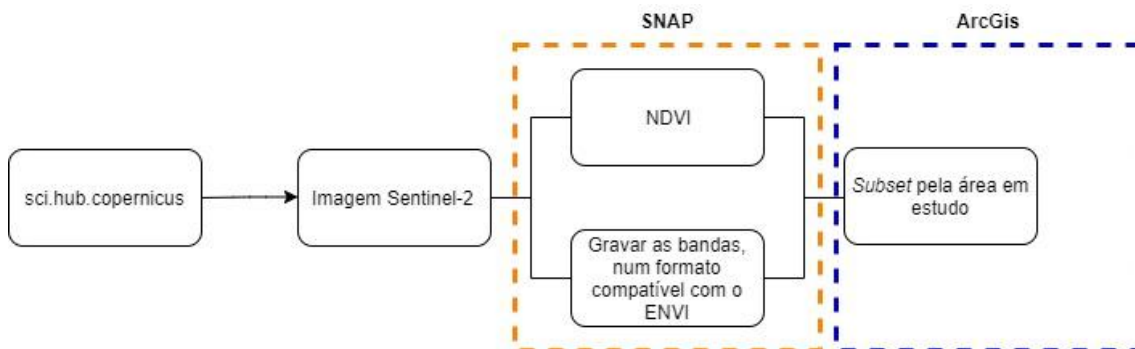


Figura 3.5. Diagrama da primeira fase do projeto

A segunda fase do projeto (Figura 3.6) consistiu na construção do mapa de intensidades dos trilhos, declives e exposição solar, todos os processos foram realizados no *software* ArcGis.

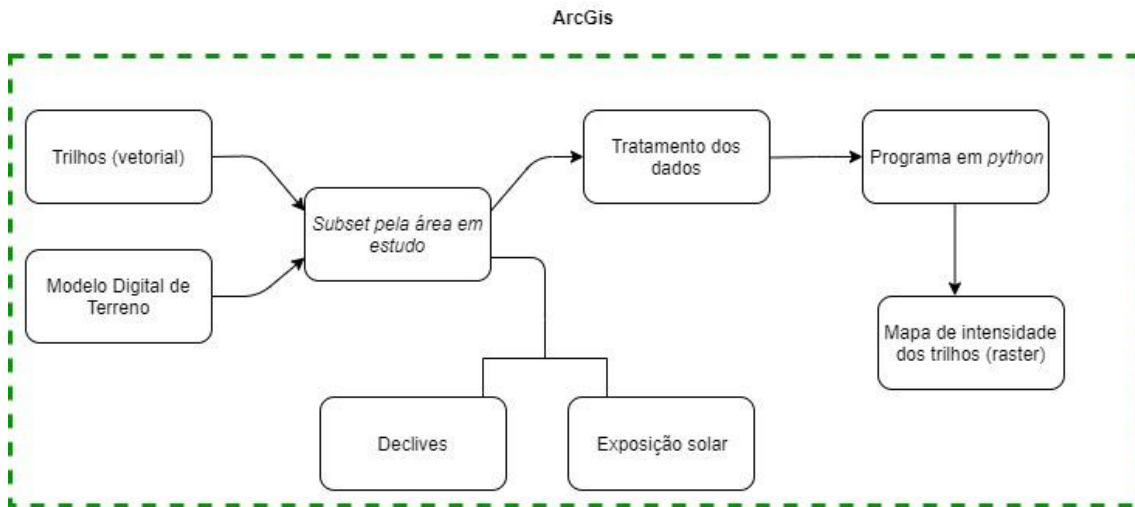


Figura 3.6. Diagrama da segunda fase do projeto

3.3.2. Classificação de Imagem

Nesta seção são descritos os algoritmos de classificação de imagem usados neste projeto.

a) Máxima Verossimilhança

O algoritmo de máxima verossimilhança é um dos mais populares e utilizados na classificação de imagens em detecção remota. É necessário determinar o valor médio da matriz da variância-covariância das respostas espectrais de cada classe, partindo do pressuposto que são amostras de uma distribuição normal multidimensional. A classificação de máxima verossimilhança é baseada na análise estatística da distribuição dos vetores espectrais da amostra para definir áreas de probabilidade equivalentes em torno desses centros. As probabilidades de cada vetor pertencer a cada classe são calculadas e o vetor é atribuído à classe para a qual obteve a probabilidade mais elevada (Richards 1986).

b) Support Vector Machine

O algoritmo *Support Vector Machine* (Figura 3.7) é responsável por encontrar *hyperplanes* que dividam os dados nas classes pretendidas, denomina-se *support vectors* os pontos mais próximos ao *hyperplane*, quanto maior a distância dos dados ao *hyperplane* maior o grau de confiança. O *hyperplane* é definido pela margem, a margem é a distância do *hyperplane* ao ponto mais próximo, o objetivo é definir um *hyperplane*

com a maior margem possível, sendo quanto maior for a margem, maior é a percentagem dos dados serem corretamente classificados.

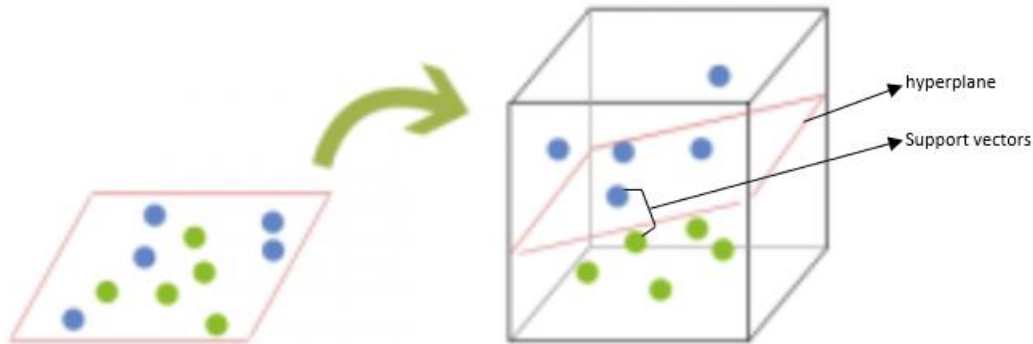


Figura 3.7. Demonstração do algoritmo SVM

c) Redes Neurais

Redes neurais são uns algoritmos específicos que vieram revolucionar a *machine learning*. São inspirados pelas redes neurais biológicas presentes no cérebro humano. As redes neurais são aproximações de funções gerais, é por esta razão que estes algoritmos podem ser aplicados a quase todos os problemas que envolvam *machine learning* (Figura 3.8).

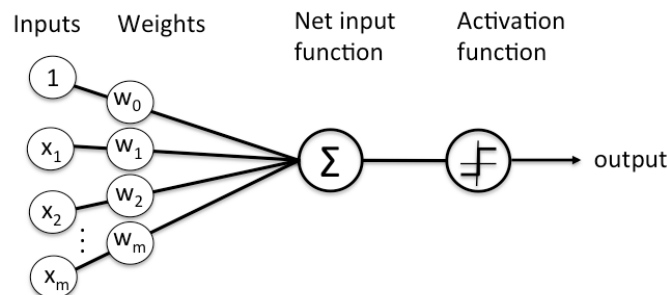


Figura 3.8 Diagrama de funcionamento de um nó da rede neuronal

d) Árvore de Decisão

Este algoritmo, é um dos mais populares utilizados em *machine learning*, e é principalmente utilizado para classificação. Uma árvore de decisão comporta-se exatamente como o cérebro humano, quando se encontra com uma questão. Uma das grandes vantagens deste algoritmo, é a facilidade de interpretação dos dados. A árvore de decisão utiliza nós que correspondem a um atributo e “folhas” que correspondem à etiqueta da classe (Figura 3.9).

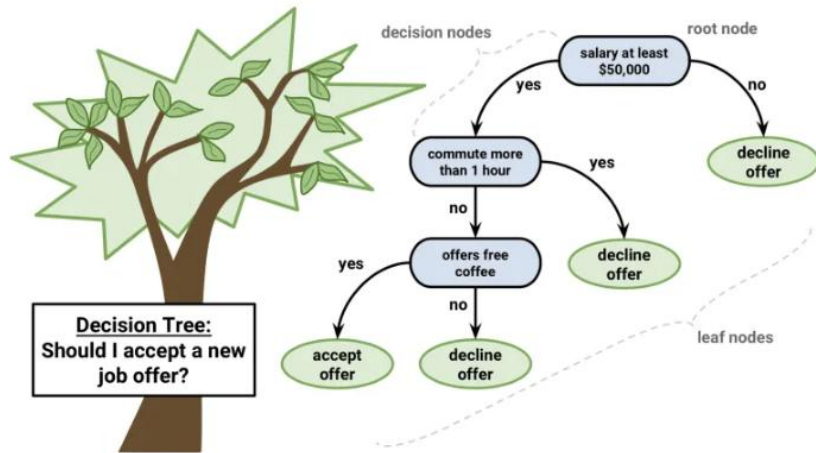


Figura 3.9. Classificação através do algoritmo Árvore de Decisão (www.psckhub.com)

Na terceira fase do projeto (figura 3.10) foi maioritariamente utilizado o *software* ENVI, onde foi efetuado a classificação das imagens com recurso aos algoritmos anteriormente mencionados. Após a classificação foi elaborada a matriz de confusão que permite avaliar a precisão de cada método utilizado para classificar a imagem. As imagens finais foram produzidas no *software* ArcGis.

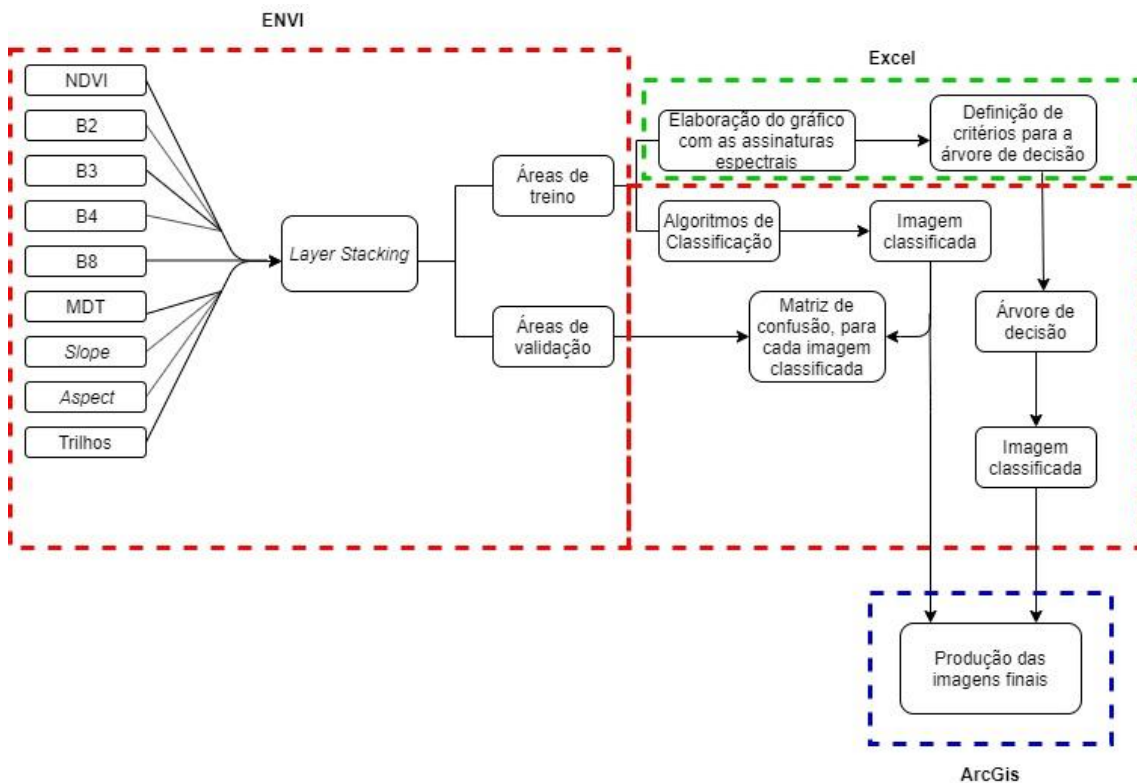


Figura 3.10 Diagrama da terceira fase do projeto

3.3.3. Validação dos Resultados Obtidos

Nesta etapa, foi utilizado única e exclusivamente o *software* Excel, na construção das tabelas com os dados provenientes das matrizes de confusão, que permitem estabelecer uma base de comparação entre os algoritmos e avaliar a sua precisão na classificação da imagem, com e sem dados auxiliares.

CAPÍTULO 4

RESULTADOS E DISCUSSÃO

Neste capítulo são apresentados os resultados do processo de classificação da imagem do Sentinel-2 usando diferentes algoritmos e com a integração e dados auxiliares provenientes dos dados geográficos voluntários.

4.1 TRATAMENTO DOS DADOS GEOGRÁFICOS VOLUNTÁRIOS

Os dados geográficos voluntários são disponibilizados em formato vetorial pelo que para que possam ser usados no processo de classificação de imagem têm de ser convertidos em formato raster. Os dados geográficos voluntários usados foram os trilhos de bicicleta e os trilhos a pé. Relativamente aos trilhos de bicicleta e a pé foram ambos cortados pela área de estudo. Foi necessário efetuar uma limpeza dos dados visto que existiam muitos dados com uma grande distância entre pontos GPS, o que inviabilizava a sua utilização (Tabela 4.1). Foi utilizado uma distância média entre pontos GPS de 40m.

Tabela 4.1 Percentagem dos dados eliminada.

Tipologia	TOTAL ENTIDADES	Entidades eliminadas	Entidades usadas	Query	% Elimina da
Bicicleta	3510	1744	1766	Trkp_dist ≤ 40m	49,69
A Pé	1380	487	893	Trkp_dist ≤ 40m	35,29
<u>Total</u>	4890	2231	2659	----- -	45,62

Os trilhos em formato vetorial foram convertidos em formato raster usando para o efeito um programa desenvolvido em *python* na consola do *software* ArcGis de modo que cada trilho fosse somado num único raster com um valor de uma unidade. O ficheiro raster final (Figura 4.1) representa a distribuição espacial dos 2659 trilhos no PNArr. A sobreposição de vários trilhos numa mesma posição resulta num aumento da intensidade da imagem. Quanto maior a intensidade maior o número de utilizadores percorreram esse trilho.

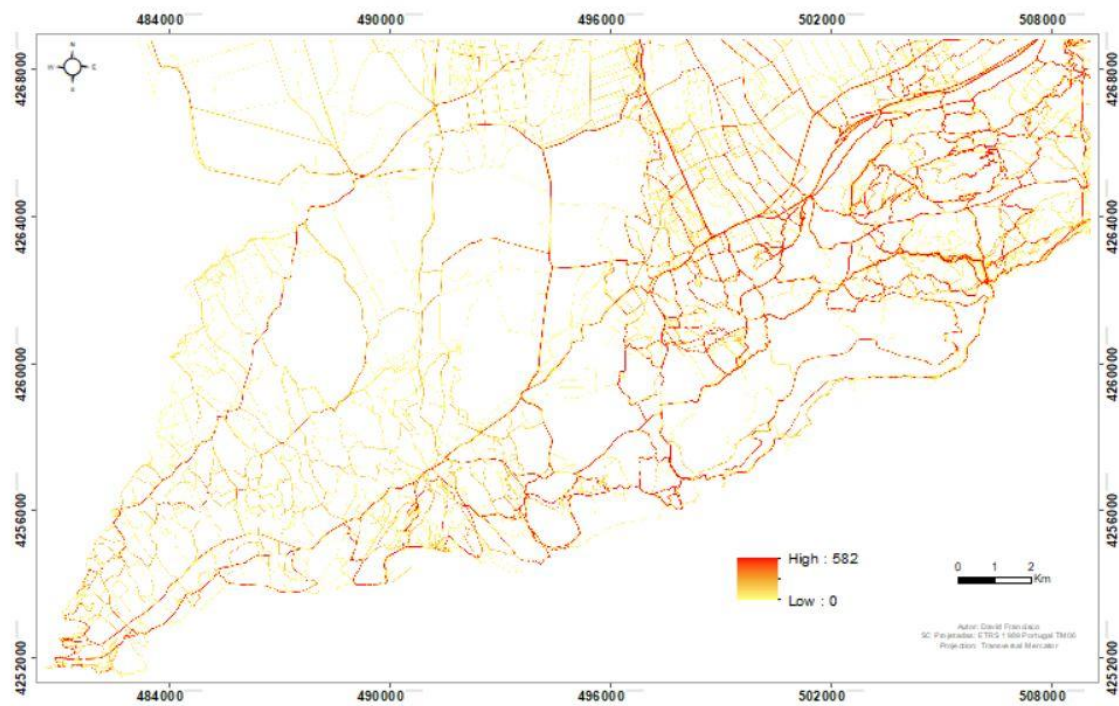


Figura 4.1. Máscara dos trilhos

Na figura 4.2, é possível observar com maior detalhe a máscara dos trilhos, na zona entre o Cabo Espichel e a vila de Sesimbra.

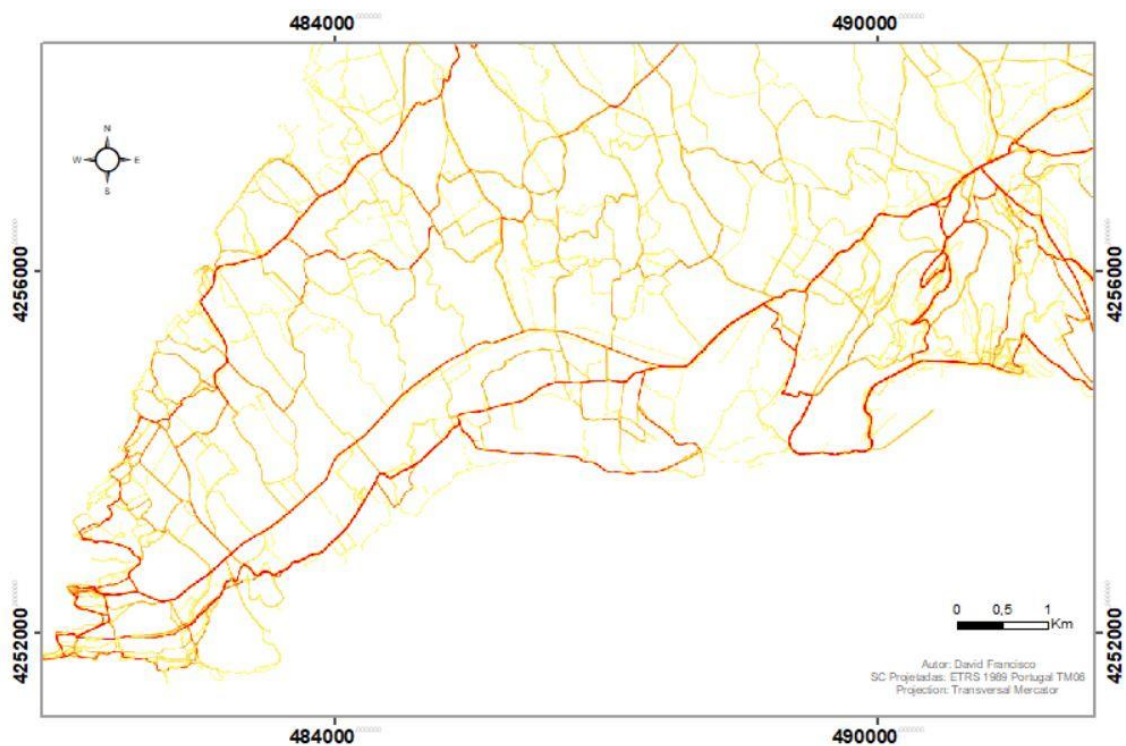


Figura 4.2 Máscara dos trilhos ampliada (Cabo Espichel/Sesimbra)

Também nesta fase, foi utilizado o modelo digital de terreno (Figura 4.3), para construir um mapa com a exposição solar e outro com os declives.

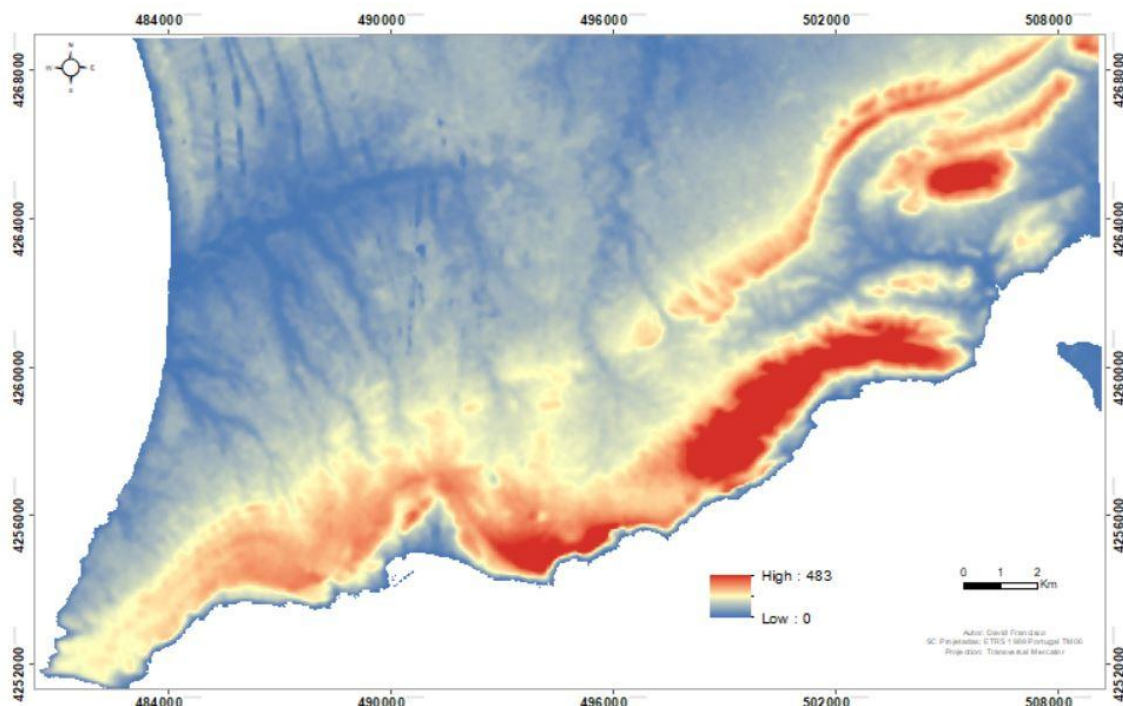


Figura 4.3. Modelo Digital de Terreno (Fonte:SNIG)

4.2 CLASSIFICAÇÃO DE IMAGEM

Nesta fase, foram utilizados três *softwares* distintos, o ENVI para o processo de classificação de imagens, o *Excel* para a construção do gráfico das assinaturas espectrais, e o ArcGis na produção das imagens finais. Foi criada uma imagem composição de nove bandas, com quatro bandas espectrais (Azul, Verde, Vermelho e Infravermelho próximo), e com cinco bandas com informação dos trilhos das bicicletas (explicada na secção anterior), do declive, da exposição solar, do modelo digital de terreno e do NDVI. Em seguida, como os algoritmos utilizados são de aprendizagem supervisionada foram criados dois conjuntos de dados, dados de treino e os dados de validação (Tabela 4.2). Os dados de treino, são utilizados para treinar os algoritmos, já os dados de validação, vão servir para efetuar a matriz de confusão para cada algoritmo, e assim avaliar a sua precisão e comparar os resultados. Foram definidas seis classes: Água, Áreas Verdes, Solo Nu, Terreno Urbanizado, Vias Artificiais, Trilhos.

Tabela 4.2. Número de pixels por classe nas áreas de treino e validação

Áreas de Treino	Número de Pixels	Áreas de Validação	Número de Pixels
Água	10873	Água	21265
Áreas Verdes	13452	Áreas Verdes	36371
Solo Nu	1664	Solo Nu	1150
Terreno Urbanizado	586	Terreno Urbanizado	1123
Vias Artificiais	541	Vias Artificiais	701
Trilhos	185	Trilhos	313

Como é possível verificar o número de pixels de classes como as da água, áreas verdes, têm um número de pixels bastante superiores às outras classes. Isto acontece, visto que a resolução espacial do sensor é de 10m, e torna-se muito mais fácil identificar classes como a água ou áreas verdes através da análise visual de imagem, e assim sendo definir os polígonos. Este maior número de pixels por parte destas duas classes vai influenciar o valor do coeficiente Kappa, um importante índice para analisarmos as matrizes de confusão. No entanto neste trabalho projeto, é importante dar mais atenção à classificação específica da classe trilhos e das vias artificiais.

Para os algoritmos da máxima verossimilhança, SVM e rede neuronal, foram apenas utilizadas as classes de treino para apoiar a sua aprendizagem. No algoritmo árvore de decisão foi construída uma “árvore” tendo por base os valores da assinatura espectral de cada classe (Figura 4.4).

Tabela 4.3. Número de cada banda no Layer Stacking

	<i>Band</i>
<i>NDVI</i>	1
<i>B2</i>	2
<i>B3</i>	3
<i>B4</i>	4
<i>B8</i>	5
<i>MDT</i>	6
<i>Slope</i>	7
<i>Aspect</i>	8
<i>Trilhos</i>	9

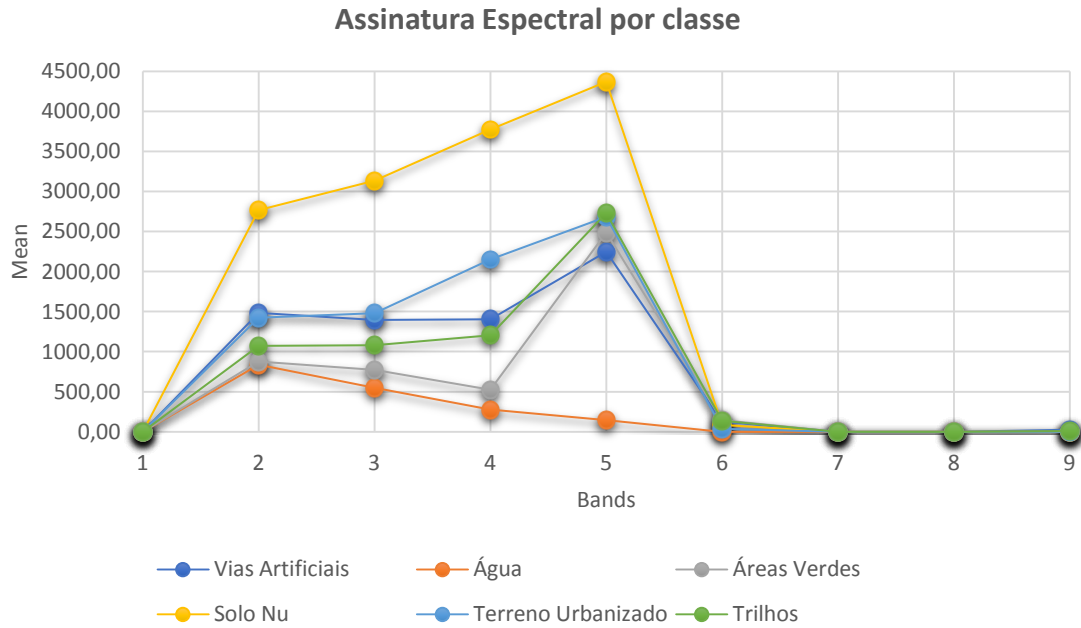


Figura 4.4. Gráfico com a assinatura espectral por classe em cada banda

Através da análise das assinaturas espectrais foram criadas regras para definir as seis classes, implementando-as numa árvore de decisão (Figura 4.5).

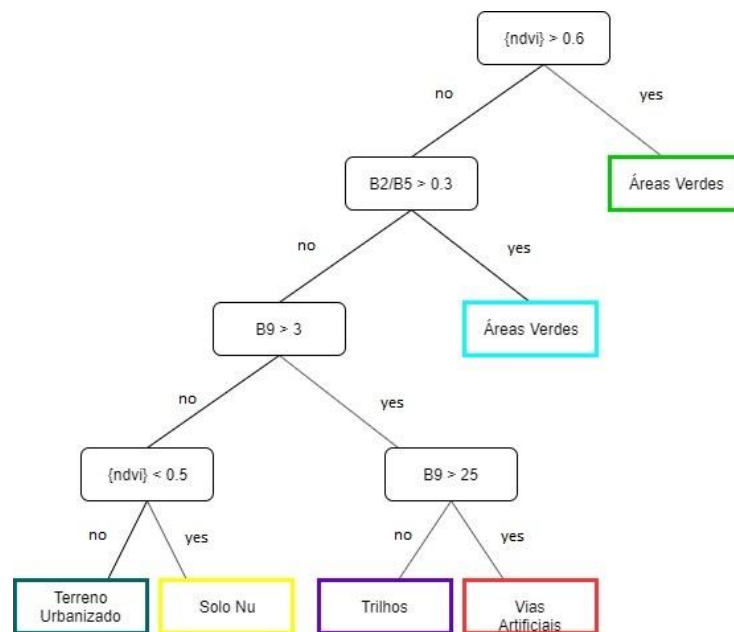


Figura 4.5. Árvore de Decisão com os critérios utilizados para a classificação da imagem

Os critérios utilizados para a classificação com a árvore de decisão, incluem alguns dados auxiliares, o que levou a que se obtivesse resultados muito melhores. Não foram utilizados todos os dados auxiliares na construção desta árvore de decisão, com o objetivo de obter o melhor resultado possível após todas as tentativas que foram feitas.

Para cada algoritmo, foram feitas duas classificações, com exceção do algoritmo árvore de decisão: uma com os dados de treino, sem qualquer dado auxiliar extra, e a outra utilizando alguns dados auxiliares como o NDVI e a máscara com os trilhos. A matriz de confusão tem vários dados que podemos utilizar para medir a precisão da classificação, no entanto, neste estudo apenas vai ser referida a tabela que contem a percentagem de pixéis bem classificados e a análise visual de imagem.

As imagens apresentadas representam uma ampliação na zona do Cabo Espichel/Sesimbra, de forma a facilitar a interpretação e observação das classes classificadas, em especial a classe dos trilhos e das vias artificiais. As imagens classificadas da totalidade da área em estudo estão integradas no anexo A.

4.3 RESULTADOS DO ALGORITMO DE MÁXIMA VEROSSIMILHANÇA

As figuras seguintes representam a classificação da imagem pelo algoritmo máxima verossimilhança (Figura 4.6 e 4.7)

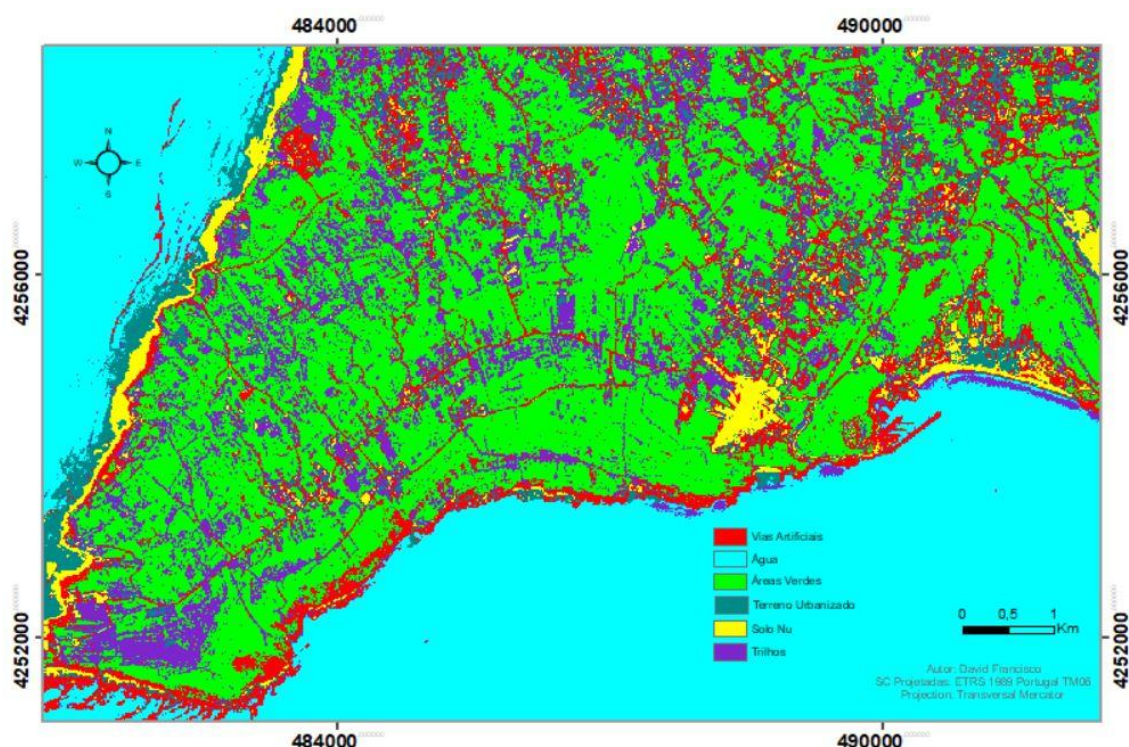


Figura 4.6. Classificação Máxima Verossimilhança sem dados auxiliares

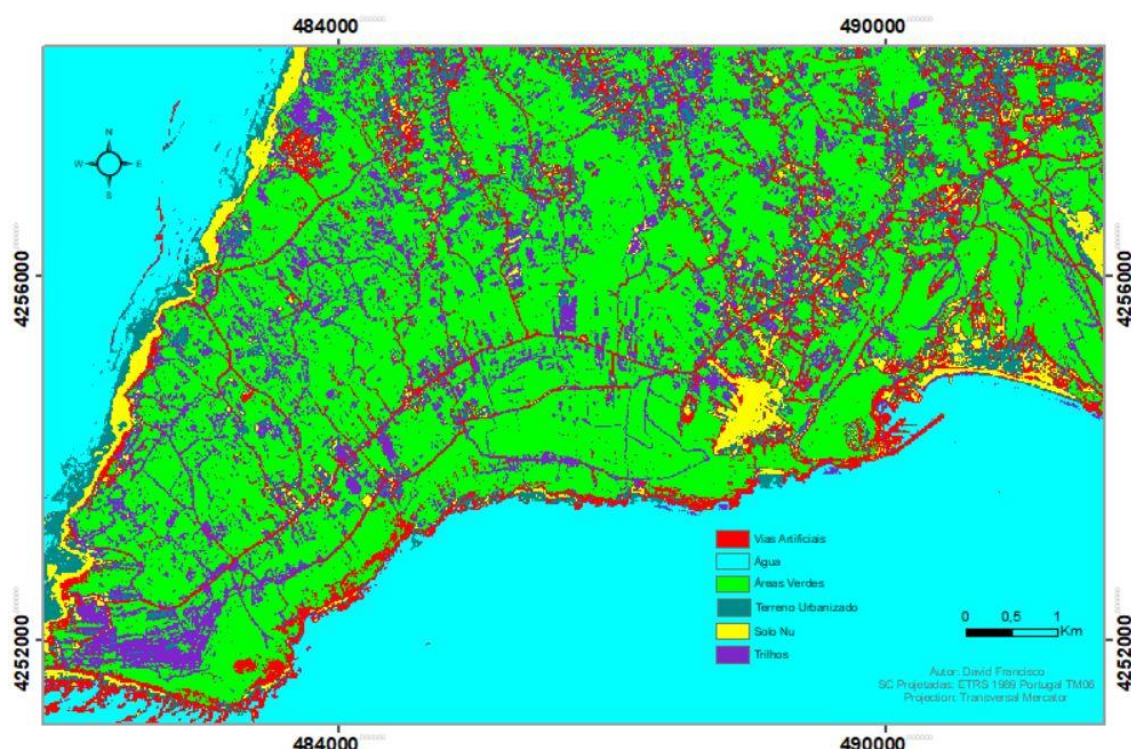


Figura 4.7. Classificação Máxima Verossimilhança com dados auxiliares

A análise visual de ambas as imagens mostra uma melhoria, não da classificação da classe trilhos, mas da redução de trilhos classificados incorretamente, na classe água. Visualmente são bastante semelhantes, não se podendo afirmar se houve ou não melhoria.

A tabela 4.4 e 4.5 representa a percentagem de pixéis bem identificados na classe certa. As colunas representam as áreas de validação e as linhas representam as áreas de treino. A primeira tabela refere-se à classificação de imagem sem dados auxiliares, enquanto a segunda tabela é referente à classificação de imagem com dados auxiliares. Sombreado a verde encontra-se o valor percentual mais relevante (Trilhos classificados como trilhos).

Tabela 4.4. Matriz de confusão para a classificação com o algoritmo de Máxima Verossimilhança sem dados auxiliares

Classe	Trilhos	Solo Nu	Terreno Urbanizado	Áreas Verdes	Água	Vias Artificiais
Trilhos	94,89	0,00	5,08	2,42	1,32	0,86
Solo Nu	0,00	100	0,27	0,00	0,00	0,00
Terreno Urbanizado	0,64	0,00	93,86	0,00	0,28	0,57
Áreas Verdes	4,47	0,00	0,00	97,47	0,00	0,29
Água	0,00	0,00	0,00	0,00	98,27	0,00
Vias Artificiais	0,00	0,00	0,80	0,12	0,06	98,29

Tabela 4.5. Matriz de confusão para a classificação com o algoritmo de Máxima Verosimilhança com dados auxiliares

<u>Classe</u>	<u>Trilhos</u>	<u>Solo Nu</u>	<u>Terreno Urbanizado</u>	<u>Áreas Verdes</u>	<u>Água</u>	<u>Vias Artificiais</u>
<i>Trilhos</i>	93,29	0,00	3,37	1,92	0,07	0,86
<i>Solo Nu</i>	0,00	99,83	0,27	0,00	0,00	0,00
<i>Terreno Urbanizado</i>	0,96	0,00	95,73	0,00	0,27	1,57
<i>Áreas Verdes</i>	5,75	0,00	0,00	98,02	0,00	0,43
<i>Água</i>	0,00	0,00	0,00	0,01	99,62	0,00
<i>Vias Artificiais</i>	0,00	0,17	0,53	0,05	0,04	97,15

A análise às matrizes de confusão, é possível verificar que não houve qualquer melhoria na classificação dos trilhos utilizando os dados auxiliares, de facto a classificação sem a utilização dos dados auxiliares obteve uma maior percentagem de exatidão, 94,89%, enquanto a classificação com o uso dos dados auxiliares, teve uma exatidão de 93,29%.

4.4. RESULTADOS DO ALGORITMO SVM

Na figura 4.8 é apresentado o resultado da classificação de imagem com o algoritmo SVM usando apenas as bandas espectrais e na figura 4.9 é apresentado o resultado usando as bandas espectrais e artificiais.

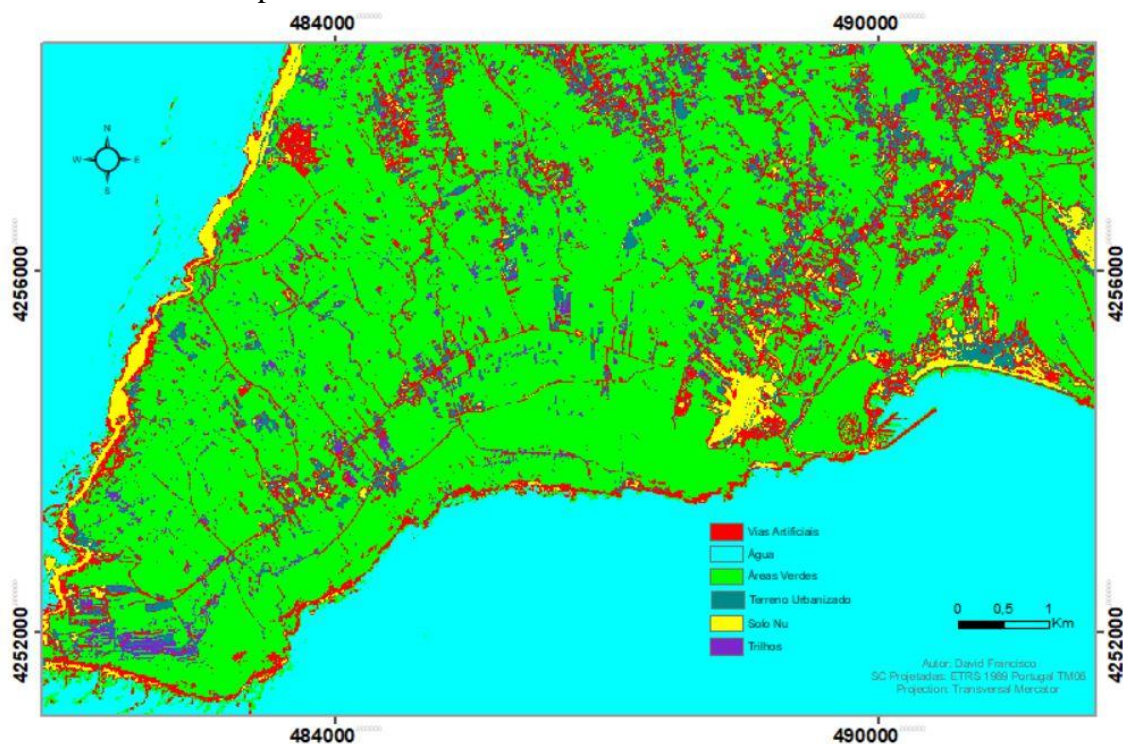


Figura 4.8. Classificação SVM sem dados auxiliares

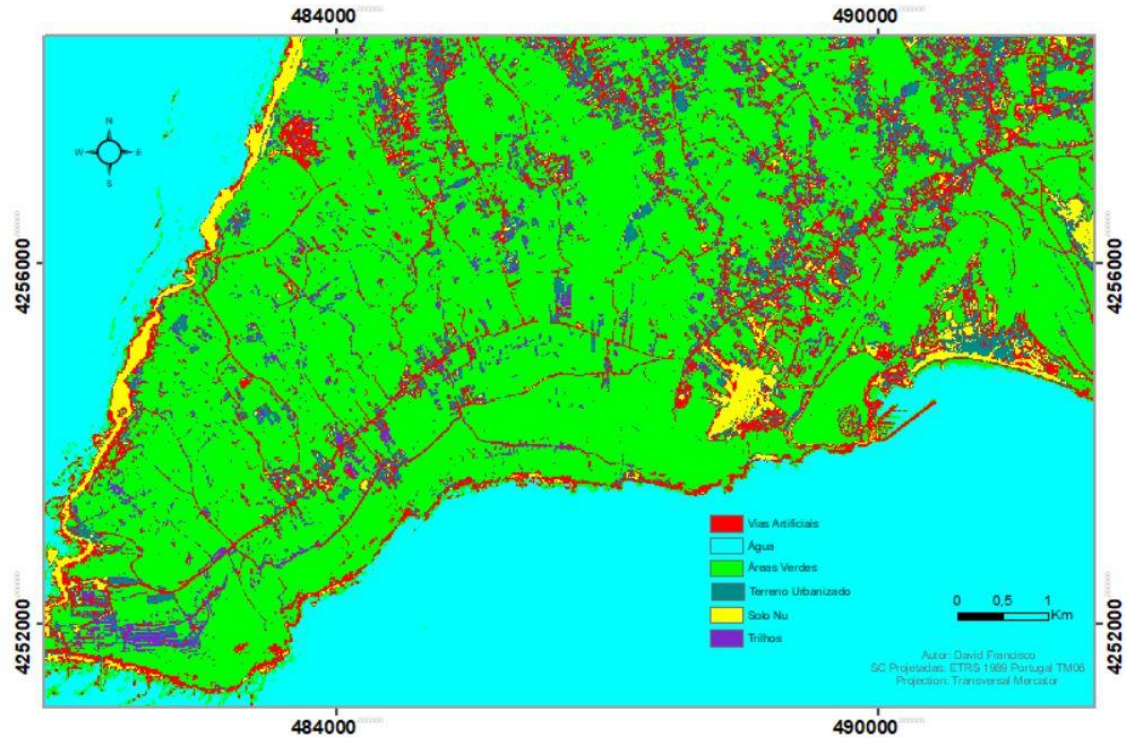


Figura 4.9. Classificação SVM com dados auxiliares

Através da análise visual de imagem às duas imagens classificadas através do método SVM, é perceptível uma melhoria na classificação da rede viária incluindo os trilhos, na imagem em que foi feita a classificação com os dados auxiliares.

A tabela 4.6 e 4.7 apresenta a percentagem de pixéis bem identificados na classe certa. As colunas representam as áreas de validação e as linhas representam as áreas de treino.

Tabela 4.6. Matriz de confusão para a classificação com o algoritmo SVM sem dados auxiliares

<u>Classe</u>	<i>Trilhos</i>	<i>Solo Nu</i>	<i>Terreno Urbanizado</i>	<i>Áreas Verdes</i>	<i>Água</i>	<i>Vias Artificiais</i>
<i>Trilhos</i>	48,56	0,00	2,40	0,02	0,00	0,43
<i>Solo Nu</i>	0,00	100	0,18	0,00	0,00	0,29
<i>Terreno Urbanizado</i>	5,11	0,00	95,99	0,00	0,32	0,29
<i>Áreas Verdes</i>	44,73	0,00	0,27	99,93	0,32	4,85
<i>Água</i>	0,00	0,00	0,00	0,05	99,67	0,00
<i>Vias Artificiais</i>	1,60	0,00	1,16	0,12	0,06	94,15

Tabela 4.7. Matriz de confusão para a classificação com o algoritmo SVM com dados auxiliares

<u>Classe</u>	<i>Trilhos</i>	<i>Solo Nu</i>	<i>Terreno Urbanizado</i>	<i>Áreas Verdes</i>	<i>Água</i>	<i>Vias Artificiais</i>
<i>Trilhos</i>	50,80	0,00	2,49	0,03	0,00	0,86
<i>Solo Nu</i>	0,00	100	0,27	0,00	0,00	0,14
<i>Terreno Urbanizado</i>	4,47	0,00	95,81	0,00	0,00	0,29
<i>Áreas Verdes</i>	42,17	0,00	0,36	99,91	0,38	3,28
<i>Água</i>	0,00	0,00	0,00	0,05	99,61	0,00
<i>Vias Artificiais</i>	2,56	0,00	1,07	0,12	0,06	95,44

Os resultados obtidos nas matrizes de confusão, corroboram a análise visual das imagens classificadas, existe de facto uma melhoria na classificação quer da classe trilhos, quer da classe vias artificiais, apesar de ligeira, pouco mais de dois pontos percentuais no que diz respeito à classe trilhos e cerca de 1% no que diz respeito à classe vias artificiais.

É de notar também que em ambas as classificações quase 50% dos trilhos foram classificados como áreas verdes.

4.5. RESULTADOS DO ALGORITMO REDE NEURONAL

As figuras seguintes representam a classificação da imagem pelo algoritmo Rede Neuronal.

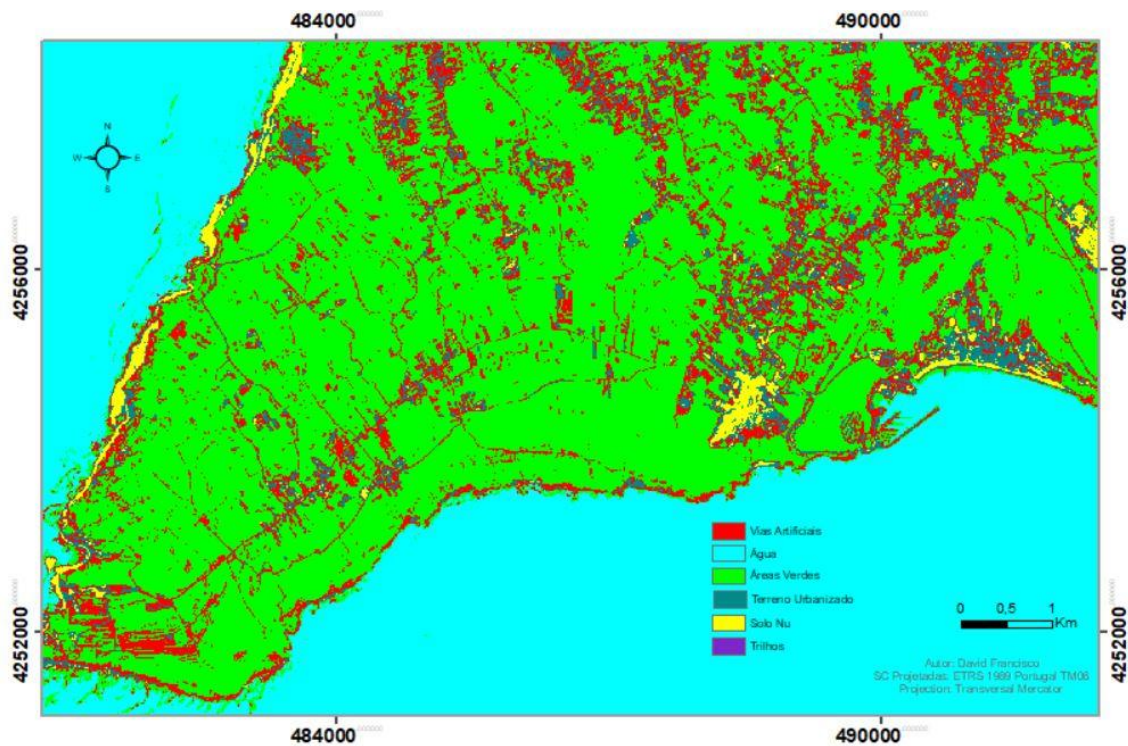


Figura 4.10. Classificação com algoritmo Rede Neuronal sem dados auxiliares

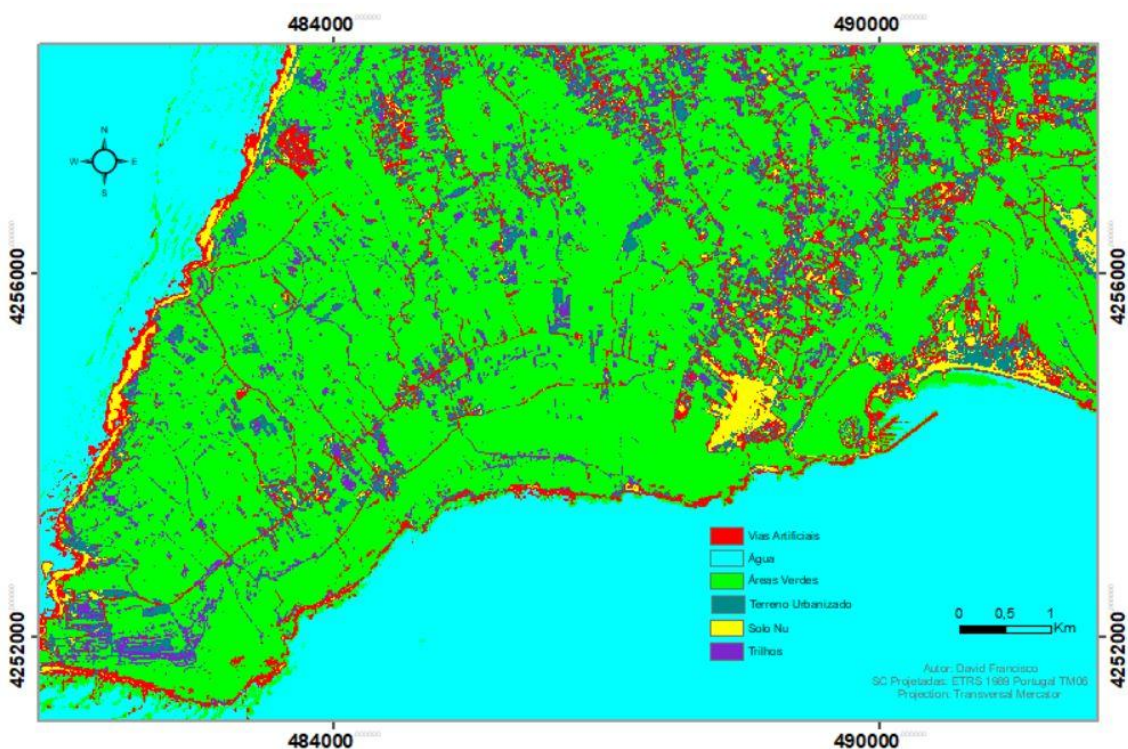


Figura 4.11. Classificação com algoritmo Rede Neuronal com dados auxiliares

As imagens classificadas demonstram uma clara diferença entre a classificação sem os dados auxiliares e classificação com os dados auxiliares. Na imagem classificada sem a utilização dos dados auxiliares, é possível verificar que os trilhos foram todos classificados como vias artificiais, já na imagem que classificada com a utilização dos dados auxiliares é possível ver que alguns trilhos foram bem classificados, ao contrário da primeira classificação.

A tabela 4.8 e 4.9 representa a percentagem de pixéis bem identificados na classe certa. As colunas representam as áreas de validação e as linhas representam as áreas de treino.

Tabela 4.8. Matriz de confusão para a classificação com o algoritmo Rede Neuronal sem dados auxiliares

<u>Classe</u>	<i>Trilhos</i>	<i>Solo Nu</i>	<i>Terreno Urbanizado</i>	<i>Áreas Verdes</i>	<i>Água</i>	<i>Vias Artificiais</i>
<i>Trilhos</i>	0,00	0,00	0,00	0,00	0,00	0,00
<i>Solo Nu</i>	0,00	98,61	1,34	0,00	0,00	0,00
<i>Terreno Urbanizado</i>	0,64	1,39	72,22	0,00	0,28	3,28
<i>Áreas Verdes</i>	65,18	0,00	0,62	99,98	0,38	7,56
<i>Água</i>	0,00	0,00	0,00	0,00	99,62	0,00
<i>Vias Artificiais</i>	34,19	0,00	25,82	0,02	0,00	89,16

Tabela 4.9. Matriz de confusão para a classificação com o algoritmo Rede Neuronal sem dados auxiliares

<u>Classe</u>	<i>Trilhos</i>	<i>Solo Nu</i>	<i>Terreno Urbanizado</i>	<i>Áreas Verdes</i>	<i>Água</i>	<i>Vias Artificiais</i>
<i>Trilhos</i>	47,92	0,00	2,94	0,05	0,00	2,57
<i>Solo Nu</i>	0,00	93,48	0,36	0,00	0,00	0,00
<i>Terreno Urbanizado</i>	4,47	6,52	95,99	0,00	0,00	1,00
<i>Áreas Verdes</i>	47,50	0,00	0,09	99,95	0,19	6,13
<i>Água</i>	0,00	0,00	0,00	0,00	99,81	0,00
<i>Vias Artificiais</i>	0,00	0,00	0,62	0,00	0,00	90,30

As análises às matrizes de confusão comprovam, o que já tinha sido possível verificar, após a análise visual às imagens classificadas, na imagem classificada sem a utilização de dados auxiliares, não houve nenhum trilho classificado como trilho, enquanto na imagem classificada com a utilização de dados auxiliares 47,92% dos trilhos foram

classificados como trilhos, o que representa uma melhoria de 47,92%, na utilização dos dados auxiliares.

4.6. RESULTADOS DO ALGORITMO ÁRVORE DE DECISÃO

Este algoritmo, dos quatro algoritmos utilizados foi o único que não foi treinado diretamente a partir das áreas de treino, foram sim, definidos critérios, através da análise ao gráfico das assinaturas espectrais de cada classe. Foram utilizados dados auxiliares, como o NDVI e a máscara dos trilhos (Figura 4.12)

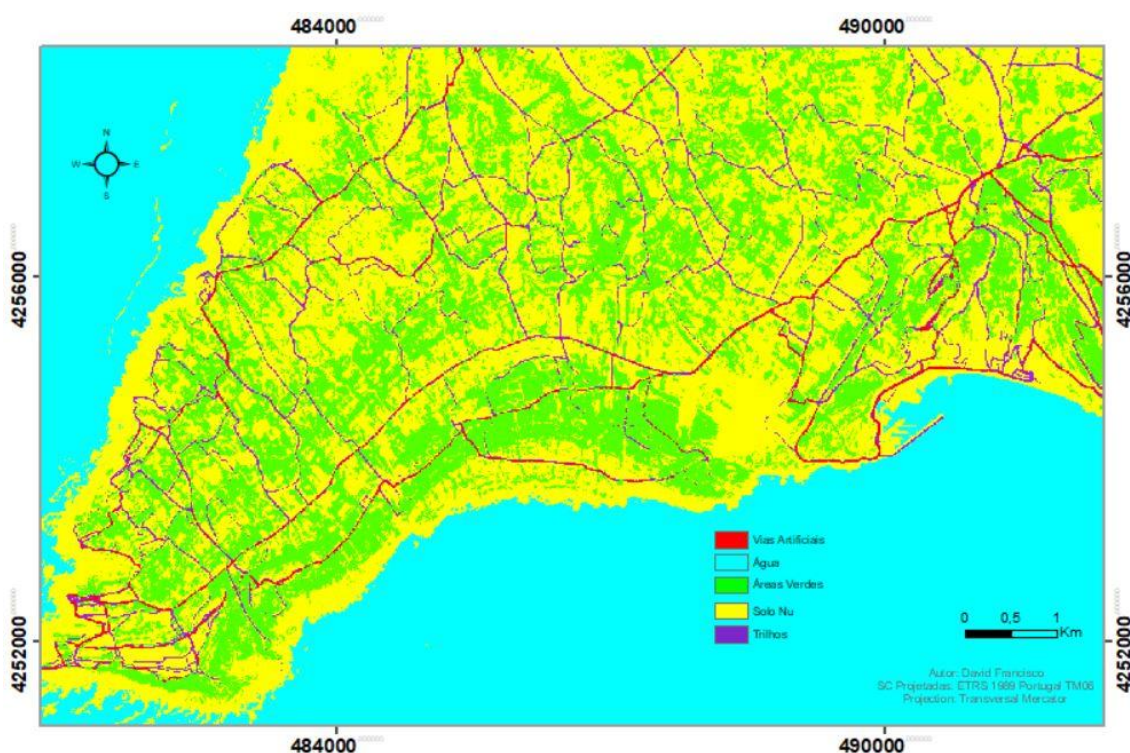


Figura 4.12. Classificação através do algoritmo de Árvore de Decisão com dados auxiliares

A análise à imagem classificada com o algoritmo árvore de decisão, revela que em comparação com as restantes imagens classificadas através dos outros algoritmos, a rede viária incluindo os trilhos está praticamente completa e bem classificada. Ao contrário das outras classificações, a base por de trás deste algoritmo está um conjunto de critérios definidos através da análise espectral das classes, análise essa que se obteve através das áreas de treino.

Este algoritmo não inclui a classe “Território Artificializado”, devido aos maus resultados obtidos nesta classe.

A Figura 4.13 mostra apenas a classificação dos trilhos e das vias artificiais.

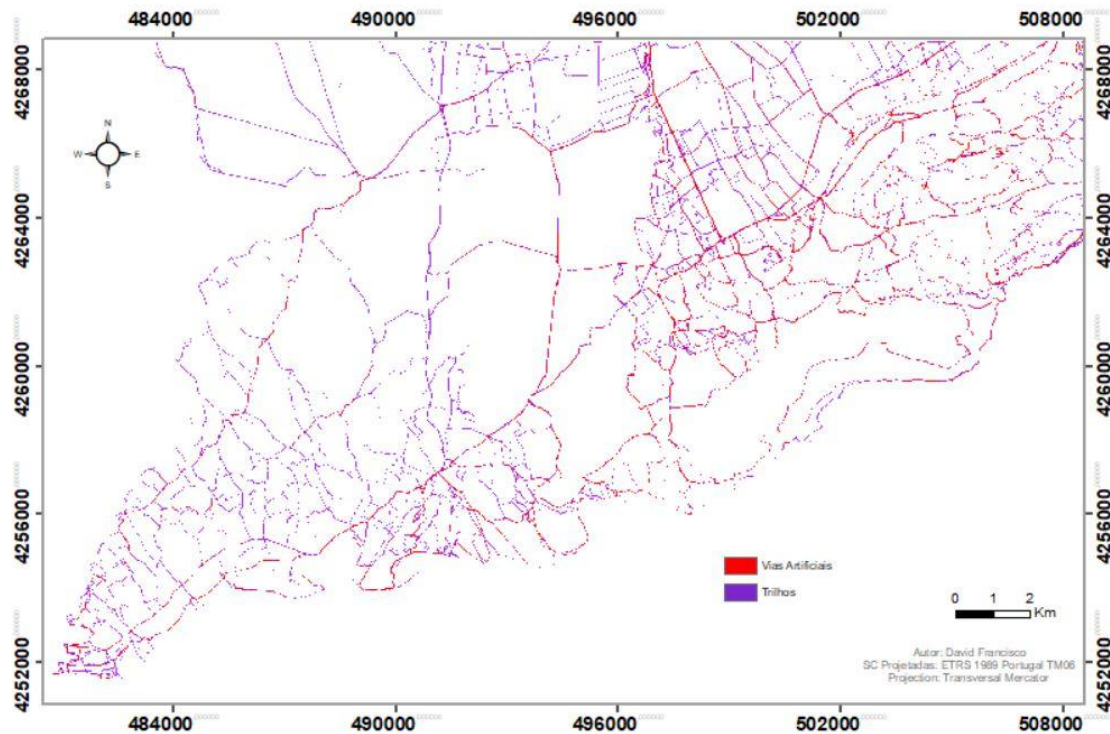


Figura 4.13. Classificação dos trilhos e das vias artificiais

Grande parte dos trilhos iniciais foram classificados por este método, como é possível observar na imagem apresentada anteriormente.

Tabela 4.10 Matriz de confusão para a classificação com o algoritmo árvore de decisão, com dados auxiliares

<u>Classe</u>	<i>Trilhos</i>	<i>Solo Nu</i>	<i>Áreas Verdes</i>	<i>Água</i>	<i>Vias Artificiais</i>
<i>Trilhos</i>	27,96	0,7	0,16	0,00	52,44
<i>Solo Nu</i>	46,24	99,13	4,38	0,12	13,32
<i>Áreas Verdes</i>	3,58	0,00	95,40	0,19	0,14
<i>Água</i>	0,00	0,00	0,00	99,88	0,00
<i>Vias Artificiais</i>	22,22	0,17	0,06	0,00	34,1

Na análise à matriz de confusão, é possível verificar que apenas 27% da classe trilhos foi efetivamente classificada como trilhos, apesar de na análise visual de imagem ser facilmente identificável grande percentagem dos trilhos existentes. Esta situação pode ser explicada pela dificuldade na criação das áreas de validação devido à resolução espacial da imagem de satélite.

No entanto através da análise visual de imagem e no que diz respeito à classe trilhos e à classe vias artificiais, este método foi o mais preciso.

4.7. COMPARAÇÃO ENTRE OS ALGORITMOS

O gráfico seguinte, representa a percentagem de classificação da classe trilhos, como trilhos. Integram o gráfico os quatro algoritmos utilizados neste trabalho de projeto.

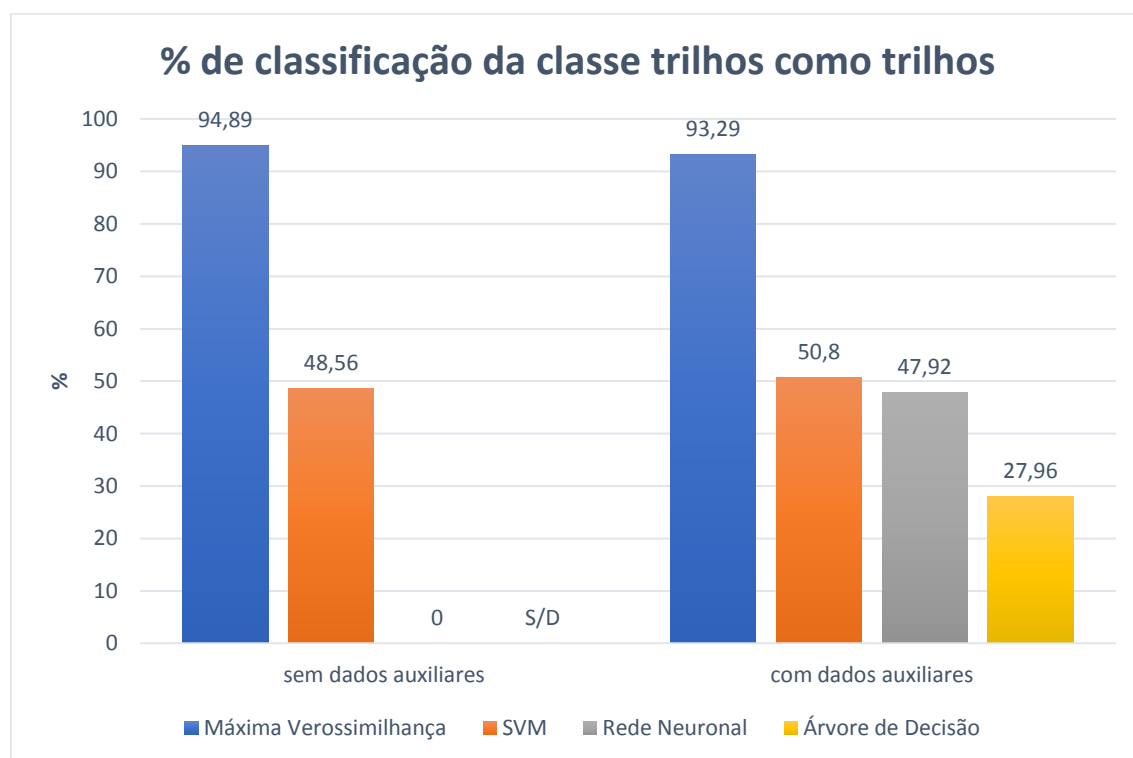


Figura 4.14. Gráfico com a percentagem da classe de trilhos classificada como trilhos por algoritmo

O algoritmo máxima verossimilhança foi o que obteve a percentagem de trilhos classificada como trilhos mais elevada, foi também o único que sofreu uma quebra na precisão da classificação de imagem sem dados auxiliares, para a classificação de imagem com dados auxiliares.

O algoritmo SVM, teve uma ligeira melhoria de cerca de 2%, mas mesmo assim a sua exatidão não foi além dos 50,8% na classificação com o recurso dados auxiliares.

A Rede Neuronal, foi o algoritmo que sofreu a maior alteração na classificação sem recurso aos dados auxiliares, para a classificação com recurso aos dados auxiliares, 47,2%, no entanto é também o método que apresenta piores resultados em ambas as classificações, visto que na primeira teve uma exatidão de 0% e na segunda classificação a sua exatidão foi de 47,2%.

O algoritmo árvore de decisão obteve uma precisão de cerca de 27%, relativamente mais baixa em relação aos outros métodos em estudo.

Em suma se tivermos em conta as análises visuais de imagem, o método árvore de decisão, visualmente apresenta com uma larga margem sobre os anteriores uma melhor classificação, classificando a maior parte dos trilhos e vias artificiais existentes no Parque Natural da Arrábida. As percentagens representadas na matriz de confusão são relativas às áreas de validação criadas manualmente.

CAPÍTULO 5

CONCLUSÃO E PERSPETIVAS FUTURAS

Este trabalho de projeto conciliou a aquisição de dados geográficos voluntários, com técnicas de deteção remota, para a construção de uma rede viária na área de estudo do Parque Natural da Arrábida, através de quatro algoritmos de classificação de imagem: Máxima Verossimilhança, *Support Vector Machine*, Rede Neuronal e Árvore de Decisão.

Os dados geográficos voluntários foram extraídos do *Open Street Map*, uma plataforma de dados geográficos voluntários, a imagem do satélite Sentinel-2 foi extraída do repositório de imagens da *European Space Agency*.

Foram elaboradas para cada um dos algoritmos, com exceção da árvore de decisão, duas classificações, uma sem recurso aos dados auxiliares, outra com recurso aos dados auxiliares, de forma a poder estabelecer uma comparação, se existe vantagem na utilização dos dados auxiliares. No geral os dados auxiliares vieram aumentar a precisão e melhorar a classificação. No caso do algoritmo árvore de decisão, a utilização dos dados auxiliares na classificação de imagem, permite praticamente identificar toda rede viária existente na área em estudo. Em relação aos outros algoritmos, e tendo em conta a precisão apresentada nas matrizes de decisão, o método de máxima verossimilhança sofreu uma quebra de precisão ao serem adicionados os dados auxiliares à classificação de imagem, tanto o SVM, como a rede neuronal, tiveram um aumento de precisão.

No que diz respeito às perspetivas futuras, há espaço para este método ser melhorado, aumentando a resolução espacial da imagem de satélite, que neste caso sendo de 10m condicionou de certa forma, principalmente na identificação das áreas de treino e validação, onde a maior parte dos trilhos têm entre dois a cinco metros de largura.

Os dados voluntários estão a evoluir cada vez mais, e cada vez há mais informação disponível e de qualidade, todos setores vão beneficiar com o aumento destes dados, que por sua vez aumenta a qualidade e a fiabilidade. No caso dos dados geográficos voluntários em específico, é uma área que tem vindo a crescer exponencialmente, estes dados são utilizados não só em projetos académicos, já são utilizados também profissionalmente. O número de utilizadores aumenta dia após dia, e foram esses dados que tornaram possível a elaboração deste trabalho de projeto

REFERÊNCIAS BIBLIOGRÁFICAS

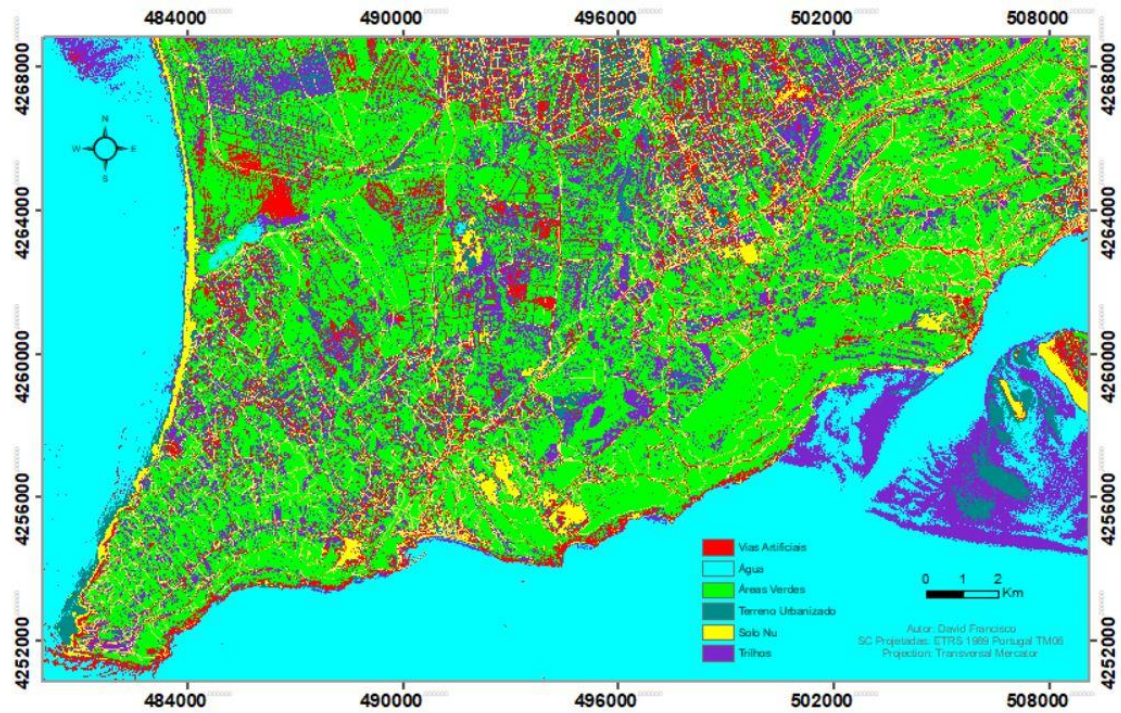
- Internet Society (2017). Artificial Intelligence and Machine Learning: Policy Paper. (April).
- Brovelli, M. A., Minghini, M., Molinari, M. E., Wu, H., Zheng, X., & Chen, J. (2018). Capacity building for high-resolution land cover intercomparison and validation: What is available and what is needed. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 42(4W8), 15–22. <https://doi.org/10.5194/isprs-archives-XLII-4-W8-15-2018>
- Decreto Lei nº622/76. (1976).
- Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Bargellini, P. (2012). Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment*, 25–36.
- Foley, J. A., DeFries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., ... Snyder, P. K. (2005). Global consequences of land use. *Science*, 309(5734), 570–574. <https://doi.org/10.1126/science.1111772>
- Goodchild, M. F. (2007). CITIZENS AS SENSORS: THE WORLD OF VOLUNTEERED GEOGRAPHY 1 Michael F. Goodchild. *GeoJournal*, 69, 211–221.
- Haklay, M. (Muki). (2008). How good is Volunteered Geographical Information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B: Planning and Design*.
- Haklay, M. (Muki), Basiouka, S., Antoniou, V., & Ather, A. (2010). How Many Volunteers Does it Take to Map an Area Well? The Validity of Linus' Law to Volunteered Geographic Information. *The Cartographic Journal*, 47(4), 315–322. <https://doi.org/10.1179/000870410X12911304958827>
- Hunter, G. J. (1999). New Tools For Handling Spatial Data Quality: Moving from Academic Concepts to Practical Reality. *URISA Journal*.
- Kounadi, O. (University of S. (2009). Assessing the Quality of OpenStreetMap Data. *MSc Geographical Information Science*.
- Linda See, Peter Mooney, Giles Foody, Lucy Bastin, Alexis Comber, Jacinto Estima, Steffen Fritz, Norman Kerle, Bin Jiang, Mari Laakso, Hai-Ying Liu, Grega

- Milćinski, Matej Nikšić, Marco Painho, Andrea Pódör, A.-M. O.-R. and M. R. (2016). Crowdsourcing, Citizen Science or Volunteered Geographic Information? The Current State of Crowdsourced Geographic Information. *International Journal of Geo-Information*.
- Mooney, P., Corcoran, P., & Winstanley, A. C. (n.d.). Towards Quality Metrics for OpenStreetMap, 514–517.
- Nie, W., Yuan, Y., Kepner, W., Jackson, M., & Erickson, C. (n.d.). Assessing Impacts of Landuse Changes on Hydrology in the Upper San Pedro Watershed
- Richards, J. A. (1986). *Remote Sensing Digital Image Analysis*. Berlin, Germany: Springer-Verlag.
- Rouse, J. W., Hass, R. H., Schell, J. A., & Deering, D. W. (1973). Monitoring vegetation systems in the great plains with ERTS. *Third Earth Resources Technology Satellite (ERTS) Symposium*, 1, 309–317. <https://doi.org/citeulike-article-id:12009708>
- Sarah, E. (Department of G. U. of W. (2009). Geographic Information Science: new geovisualization technologies — emerging questions and linkages with GIScience research. *Progress in Human Geograpy*.
- Stehman, S. V., Fonte, C. C., Foody, G. M., & See, L. (2018). Using volunteered geographic information (VGI) in design-based statistical inference for area estimation and accuracy assessment of land cover. *Remote Sensing of Environment*, 212, 47–59
- Webber, P., & Haklay, M. (Muki). (2008). OpenStreetMap: User-Generated Street Maps. *IEEE Pervasive Computing*.

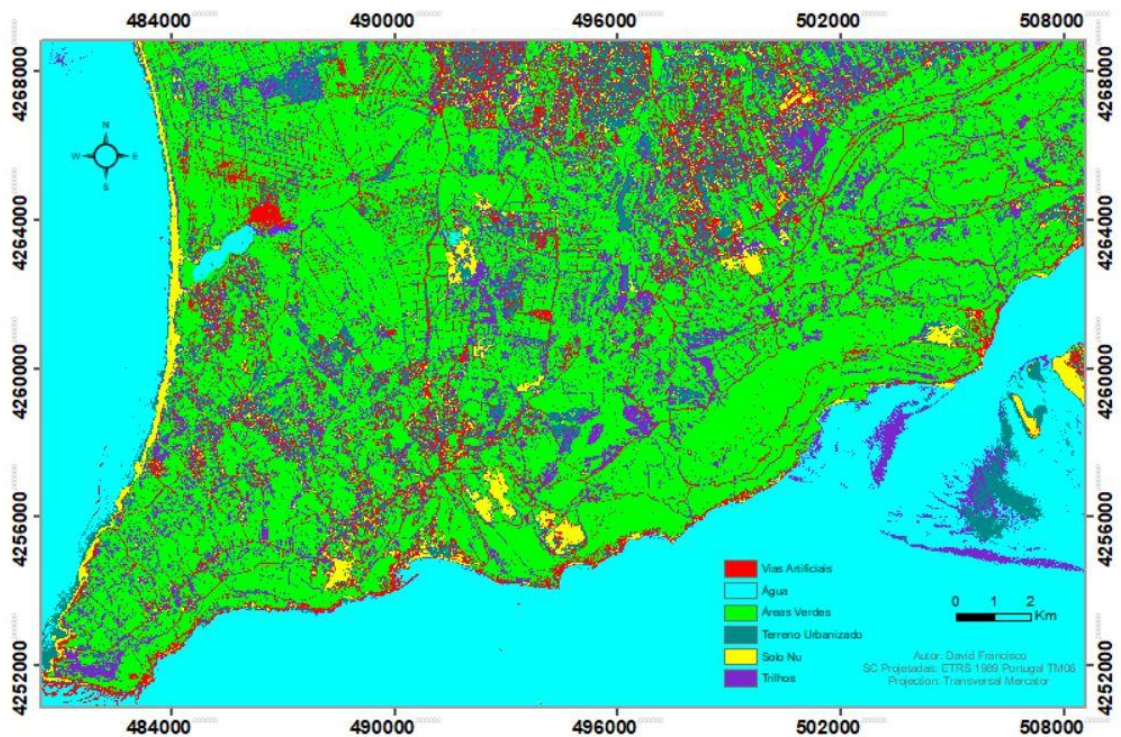
ANEXOS

A. Imagens classificadas

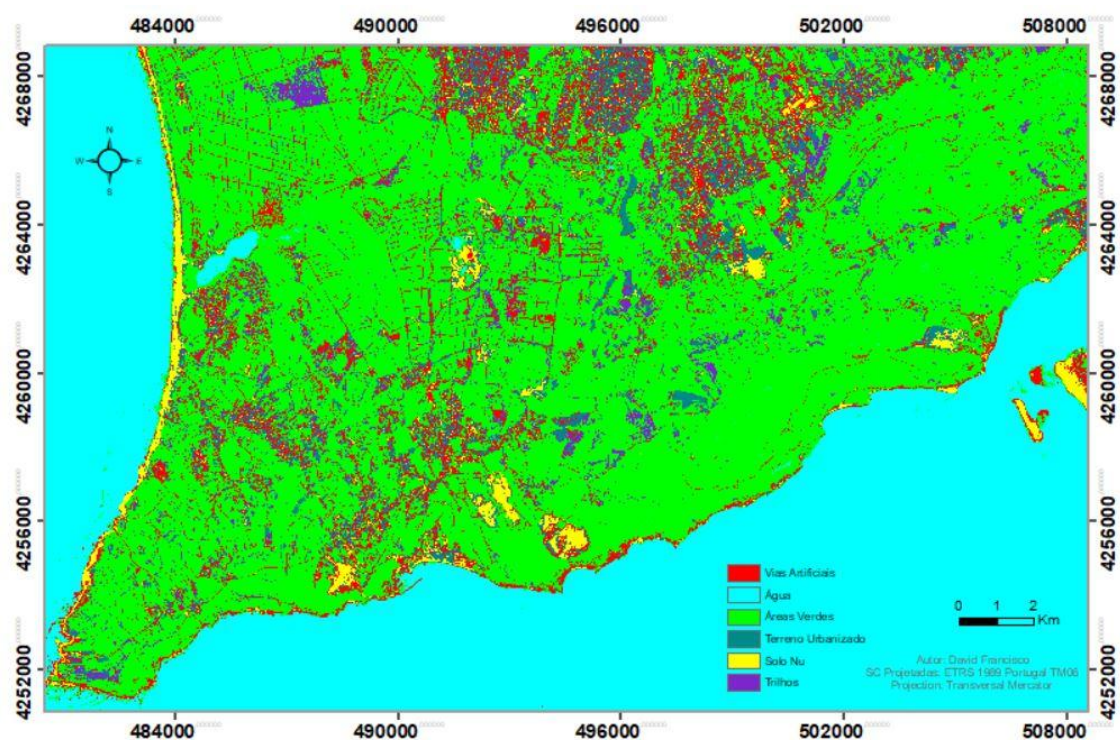
- i. Classificação através do algoritmo máxima verossimilhança sem dados auxiliares



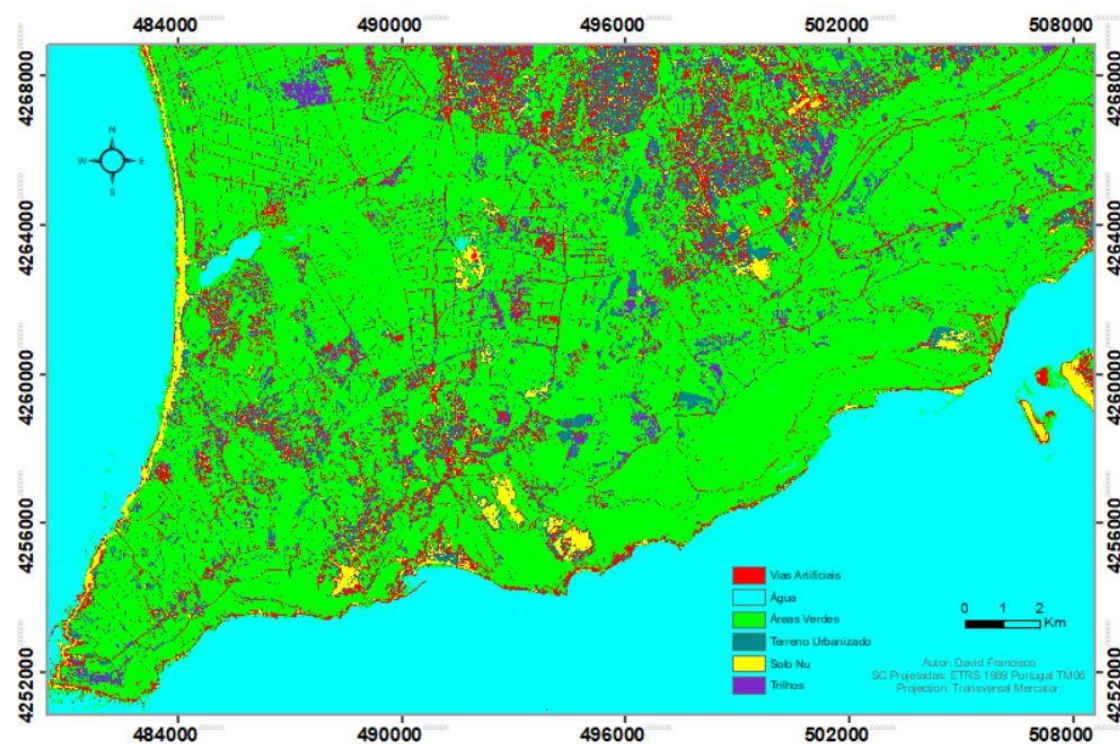
- ii. Classificação através do algoritmo máxima verossimilhança com dados auxiliares



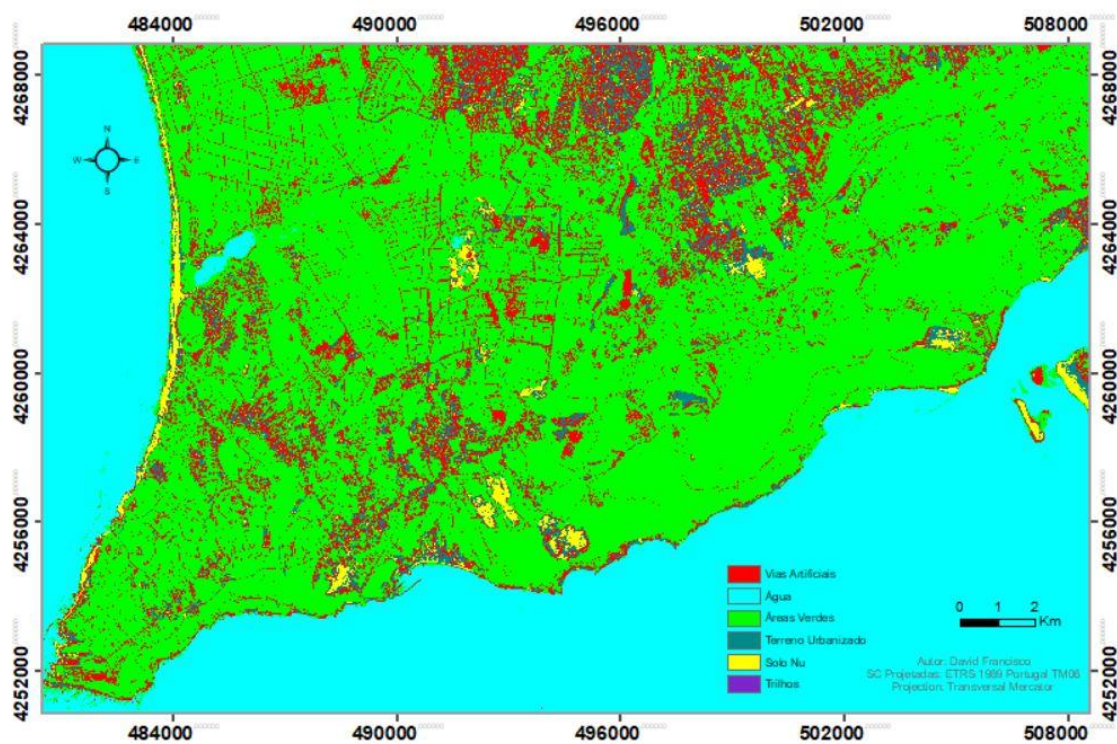
iii. Classificação através do algoritmo SVM sem dados auxiliares



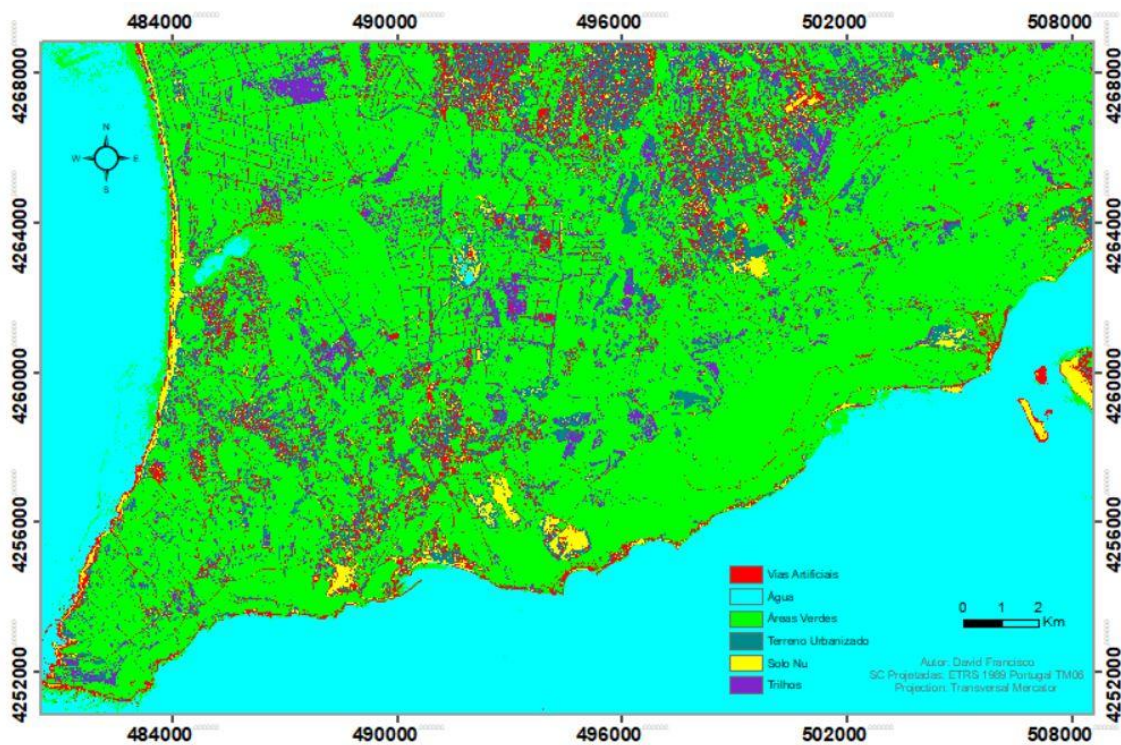
iv. Classificação através do algoritmo SVM com dados auxiliares



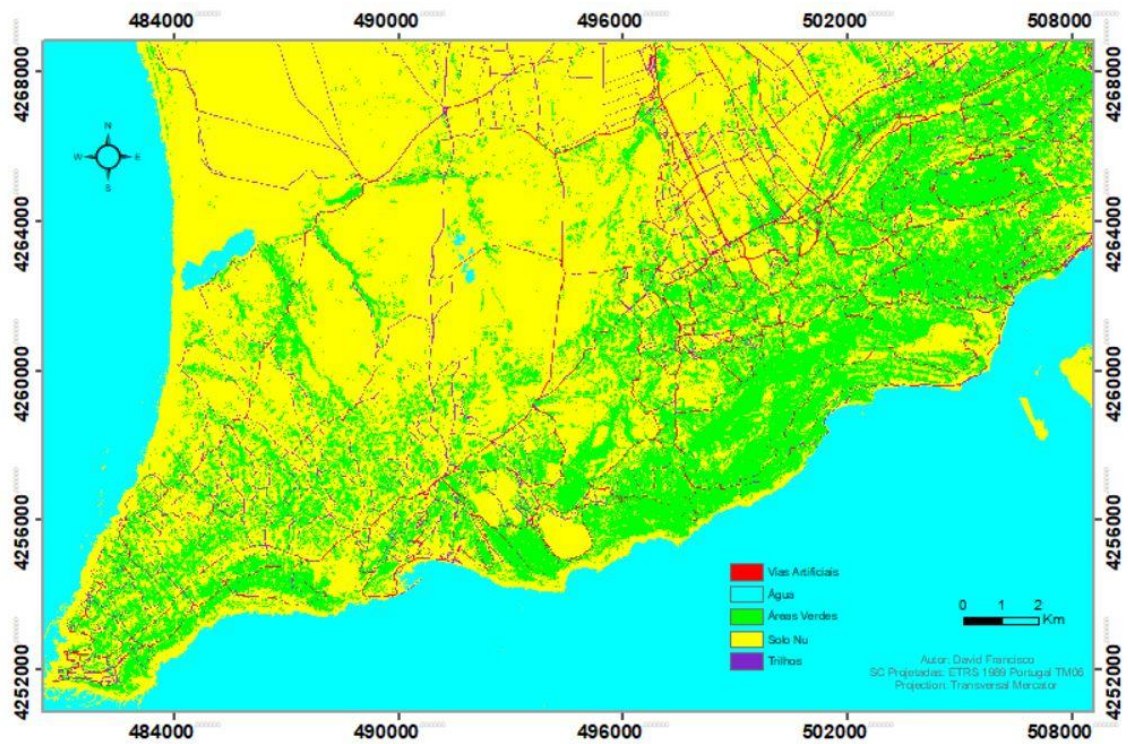
v. Classificação através do algoritmo rede neuronal sem dados auxiliares



vi. Classificação através do algoritmo rede neuronal com dados auxiliares



vii. Classificação através do algoritmo árvore de decisão com dados auxiliares



B. Matrizes de Confusão Completas

i. Matriz de confusão máxima verosimilhança sem dados auxiliares

Overall Accuracy = (59536/60923) 97.7234%

Kappa Coefficient = 0.9572

Ground Truth (Pixels)

Class	Trilhos V	Vias Artifi V	Água V	Solo Nu V	Áreas Ver T	Terreno Urban	Total
Unclassified	0	0	0	0	0	0	0
Trilhos T	297	6	280	0	880	57	1520
Vias Artifi T	0	689	13	0	42	9	753
Água T	0	0	20897	0	0	0	20987
Solo Nu T	0	0	0	1150	0	3	1153
Áreas Verd T	14	2	16	0	35449	0	35481
Terreno Urba T	2	4	59	0	0	1054	1119
Total	313	701	21265	1150	36371	1123	60923

Ground Truth (Percent)

Class	Trilhos V	Vias Artifi V	Água V	Solo Nu V	Áreas Verd T	Terreno Urban	Total
Unclassified	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Trilhos T	94.89	0.86	1.32	0.00	2.42	5.08	2.49
Vias Artifici	0.00	98.29	0.06	0.00	0.12	0.80	1.24
Água T	0.00	0.00	98.27	0.00	0.00	0.00	34.30
Solo Nu T	0.00	0.00	0.00	100.00	0.00	0.27	1.89
Áreas Verdes	4.47	0.29	0.08	0.00	97.47	0.00	58.24
Terreno Urban	0.64	0.57	0.28	0.00	0.00	93.86	1.84
Total	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Class	Commission (Percent)	Omission (Percent)	Commission (Pixels)	Omission (Pixels)
Trilhos T	80.46	5.11	1223/1520	16/313
Vias Artifici	8.50	1.71	64/753	12/701
Água T	0.00	1.73	0/20897	368/21265
Solo Nu T	0.26	0.00	3/1153	0/1150
Áreas Verdes	0.09	2.53	32/35481	922/36371
Terreno Urban	5.81	6.14	65/1119	69/1123

Class	Prod. Acc. (Percent)	User Acc. (Percent)	Prod. Acc. (Pixels)	User Acc. (Pixels)
Trilhos T	94.89	19.54	297/313	297/1520
Vias Artifici	98.29	91.50	689/701	689/753

Água T	98.27	100.00	20897/21265	20897/20897
Solo Nu T	100.00	99.74	1150/1150	1150/1153
Áreas Verdes	97.47	99.91	35449/36371	35449/35481
Terreno Urban	93.86	94.19	1054/1123	1054/1119

ii. Matriz de confusão máxima verosimilhança com dados auxiliares

Overall Accuracy = (60031/60923) 98.5359%

Kappa Coefficient = 0.9723

Ground Truth (Pixels)

Class	Trilhos V	Solo Nu V	Terreno Urban	Água V	Vias Artifici	Total
ÁreasVerdes						
Unclassified	0	0	0	0	0	0
Trilhos T	292	0	39	14	697	1048
Solo Nu T	0	1148	3	0	0	1151
Terreno Urban	3	0	1075	57	1	1147
Água T	0	0	0	21184	5	21189
Áreas Verdes	18	0	0	1	35651	3
Vias Artifici	0	2	6	9	17	681
Total	313	1150	1123	21265	36371	701
						60923

Ground Truth (Percent)

Class	Trilhos V	Solo Nu V	Terreno Urban	Água V	Vias Artifici	Total
ÁreasVerdes						
Unclassified	0.00	0.00	0.00	0.00	0.00	0.00
Trilhos T	93.29	0.00	3.47	0.07	1.92	0.86
Solo Nu T	0.00	99.83	0.27	0.00	0.00	0.00
Terreno Urban	0.96	0.00	95.73	0.27	0.00	1.57
Água T	0.00	0.00	0.00	99.62	0.01	0.00
Áreas Verdes	5.75	0.00	0.00	0.00	98.02	0.43
Vias Artifici	0.00	0.17	0.53	0.04	0.05	97.15
Total	100.00	100.00	100.00	100.00	100.00	100.00

Class	Commission (Percent)	Omission (Percent)	Commission (Pixels)	Omission (Pixels)
Trilhos T	72.14	6.71	756/1048	21/313
Solo Nu T	0.26	0.17	3/1151	2/1150
Terreno Urban	6.28	4.27	72/1147	48/1123
Água T	0.02	0.38	5/21189	81/21265
Áreas Verdes	0.06	1.98	22/35673	720/36371
Vias Artifici	4.76	2.85	34/715	20/701

Class	Prod. Acc. (Percent)	User Acc. (Percent)	Prod. Acc. (Pixels)	User Acc. (Pixels)
Trilhos T	93.29	27.86	292/313	292/1048
Solo Nu T	99.83	99.74	1148/1150	1148/1151

Terreno Urban	95.73	93.72	1075/1123	1075/1147
Água T	99.62	99.98	21184/21265	21184/21189
Áreas Verdes	98.02	99.94	35651/36371	35651/35673
Vias Artifici	97.15	95.24	681/701	681/715

iii. Matriz de confusão SVM sem dados auxiliares

Overall Accuracy = (60581/60923) 99.4386%

Kappa Coefficient = 0.9892

Ground Truth (Pixels)

Class	Trilhos V	Solo Nu V	Terreno Urban	Áreas Verdes	Água V	Vias Artifici	Total
Unclassified	0	0	0	0	0	0	0
Trilhos T	152	0	27	6	0	3	188
Solo Nu T	0	1150	2	0	0	2	1154
Terreno Urban	16	0	1078	0	0	2	1096
Áreas Verdes	140	0	3	36346	69	34	36592
Água T	0	0	0	18	21195	0	21213
Vias Artifici	5	0	13	1	1	660	680
Total	313	1150	1123	36371	21265	701	60923

Ground Truth (Percent)

Class	Trilhos V	Solo Nu V	Terreno Urban	Áreas Verdes	Vias Artifici	Total
Água V						
Unclassified	0.00	0.00	0.00	0.00	0.00	0.00
Trilhos T	48.56	0.00	2.40	0.02	0.00	0.31
Solo Nu T	0.00	100.00	0.18	0.00	0.00	1.89
Terreno Urban	5.11	0.00	95.99	0.00	0.00	1.80
Áreas Verdes	44.73	0.00	0.27	99.93	0.32	60.06
Água T	0.00	0.00	0.00	0.05	99.67	34.82
Vias Artifici	1.60	0.00	1.16	0.00	0.00	1.12
Total	100.00	100.00	100.00	100.00	100.00	100.00

Class	Commission (Percent)	Omission (Percent)	Commission (Pixels)	Omission (Pixels)
Trilhos T	19.15	51.44	36/188	161/313
Solo Nu T	0.35	0.00	4/1154	0/1150
Terreno Urban	1.64	4.01	18/1096	45/1123
Áreas Verdes	0.67	0.07	246/36592	25/36371
Água T	0.08	0.33	18/21213	70/21265
Vias Artifici	2.94	5.85	20/680	41/701

Class	Prod. Acc. (Percent)	User Acc. (Percent)	Prod. Acc. (Pixels)	User Acc. (Pixels)
-------	-------------------------	------------------------	------------------------	-----------------------

Trilhos T	48.56	80.85	152/313	152/188
Solo Nu T	100.00	99.65	1150/1150	1150/1154
Terreno Urban	95.99	98.36	1078/1123	1078/1096
Áreas Verdes	99.93	99.33	36346/36371	36346/36592
Água T	99.67	99.92	21195/21265	21195/21213
Vias Artifici	94.15	97.06	660/701	660/680

iv. Matriz de confusão SVM com dados auxiliares

Overall Accuracy = (60577/60923) 99.4321%

Kappa Coefficient = 0.9891

Ground Truth (Pixels)

Class	Trilhos V	Solo Nu V	Terreno Urban	Áreas Verdes	Água V	Vias Artifi	Total
Unclassified	0	0	0	0	0	0	0
Trilhos T	159	0	28	12	0	6	205
Solo Nu T	0	1150	3	0	0	1	1154
Terreno Urban	14	0	1076	0	0	2	1092
Áreas Verdes	132	0	4	36340	81	23	36580
Água T	0	0	0	18	21183	0	21201
Vias Artifici	8	0	12	1	1	669	691
Total	313	1150	1123	36371	21265	701	60923

Ground Truth (Percent)

Class	Trilhos V	Solo Nu V	Terreno Urban	Áreas Verdes	Água V	Vias Artifi	Total
Unclassified	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Trilhos T	50.80	0.00	2.49	0.03	0.00	0.86	0.34
Solo Nu T	0.00	100.00	0.27	0.00	0.00	0.14	1.89
Terreno Urban	4.47	0.00	95.81	0.00	0.00	0.29	1.79
Áreas Verdes	42.17	0.00	0.36	99.91	0.38	3.28	60.04
Água T	0.00	0.00	0.00	0.05	99.61	0.00	34.80
Vias Artifici	2.56	0.00	1.07	0.00	0.00	95.44	1.13
Total	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Class	Commission (Percent)	Omission (Percent)	Commission (Pixels)	Omission (Pixels)
Trilhos T	22.44	49.20	46/205	154/313
Solo Nu T	0.35	0.00	4/1154	0/1150
Terreno Urban	1.47	4.19	16/1092	47/1123
Áreas Verdes	0.66	0.09	240/36580	31/36371
Água T	0.08	0.39	18/21201	82/21265
Vias Artifici	3.18	4.56	22/691	32/701

Class	Prod. Acc. (Percent)	User Acc. (Percent)	Prod. Acc. (Pixels)	User Acc. (Pixels)
-------	-------------------------	------------------------	------------------------	-----------------------

Trilhos T	50.80	77.56	159/313	159/205
Solo Nu T	100.00	99.65	1150/1150	1150/1154
Terreno Urban	95.81	98.53	1076/1123	1076/1092
Áreas Verdes	99.91	99.34	36340/36371	36340/36580
Água T	99.61	99.92	21183/21265	21183/21201
Vias Artifici	95.44	96.82	669/701	669/691

v. Matriz de confusão rede neuronal sem dados auxiliares

Overall Accuracy = (60117/60923) 98.6770%

Kappa Coefficient = 0.9745

Ground Truth (Pixels)

Class	Trilhos V	Solo Nu V	Terreno Urban	Áreas Verdes	Água V	Vias Artifi	Total
Unclassified	0	0	0	0	0	0	0
Trilhos T	0	0	0	0	0	0	0
Solo Nu T	0	1134	15	0	0	0	1149
Terreno Urban	2	16	811	0	0	23	852
Áreas Verdes	204	0	7	36363	81	53	36708
Água T	0	0	0	0	21184	0	21184
Vias Artifici	107	0	290	8	0	625	1030
Total	313	1150	1123	36371	21265	701	60923

Ground Truth (Percent)

Class	Trilhos V	Solo Nu V	Terreno Urban	Áreas Verdes	Água V	Vias Artifi	Total
Unclassified	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Trilhos T	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Solo Nu T	0.00	98.61	1.34	0.00	0.00	0.00	1.89
Terreno Urban	0.64	1.39	72.22	0.00	0.00	3.28	1.40
Áreas Verdes	65.18	0.00	0.62	99.98	0.38	7.56	60.25
Água T	0.00	0.00	0.00	0.00	99.62	0.00	34.77
Vias Artifici	34.19	0.00	25.82	0.02	0.00	89.16	1.69
Total	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Class	Commission (Percent)	Omission (Percent)	Commission (Pixels)	Omission (Pixels)
Trilhos T	0.00	100.00	0/0	313/313
Solo Nu T	1.31	1.39	15/1149	16/1150
Terreno Urban	4.81	27.78	41/852	312/1123
Áreas Verdes	0.94	0.02	345/36708	8/36371
Água T	0.00	0.38	0/21184	81/21265
Vias Artifici	39.32	10.84	405/1030	76/701

Class	Prod. Acc. (Percent)	User Acc. (Percent)	Prod. Acc. (Pixels)	User Acc. (Pixels)
-------	-------------------------	------------------------	------------------------	-----------------------

Trilhos T	0.00	0.00	0/313	0/0
Solo Nu T	98.61	98.69	1134/1150	1134/1149
Terreno Urban	72.22	95.19	811/1123	811/852
Áreas Verdes	99.98	99.06	36363/36371	36363/36708
Água T	99.62	100.00	21184/21265	21184/21184
Vias Artifici	89.16	60.68	625/701	625/1030

vi. Matriz de confusão rede neuronal com dados auxiliares

Overall Accuracy = (60513/60923) 99.3270%

Kappa Coefficient = 0.9870

Ground Truth (Pixels)

Class	Trilhos V	Solo Nu V	Terreno Urban	Áreas Verdes	Água V	Vias Artifi	Total
Unclassified	0	0	0	0	0	0	0
Trilhos T	150	0	33	18	0	18	219
Solo Nu T	0	1075	4	0	0	0	1079
Terreno Urban	14	75	1078	0	0	7	1174
Áreas Verdes	149	0	1	36353	41	43	36587
Água T	0	0	0	0	21224	0	21224
Vias Artifici	0	0	7	0	0	633	640
Total	313	1150	1123	36371	21265	701	60923

Ground Truth (Percent)

Class	Trilhos V	Solo Nu V	Terreno Urban	Áreas Verdes	Água V	Vias Artifi	Total
Unclassified	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Trilhos T	47.92	0.00	2.94	0.05	0.00	2.57	0.36
Solo Nu T	0.00	93.48	0.36	0.00	0.00	0.00	1.77
Terreno Urban	4.47	6.52	95.99	0.00	0.00	1.00	1.93
Áreas Verdes	47.60	0.00	0.09	99.95	0.19	6.13	60.05
Água T	0.00	0.00	0.00	0.00	99.81	0.00	34.84
Vias Artifici	0.00	0.00	0.62	0.00	0.00	90.30	1.05
Total	100.00	100.00	100.00	100.00	100.00	100.00	100.00

Class	Commission (Percent)	Omission (Percent)	Commission (Pixels)	Omission (Pixels)
Trilhos T	31.51	52.08	69/219	163/313
Solo Nu T	0.37	6.52	4/1079	75/1150
Terreno Urban	8.18	4.01	96/1174	45/1123
Áreas Verdes	0.64	0.05	234/36587	18/36371
Água T	0.00	0.19	0/21224	41/21265
Vias Artifici	1.09	9.70	7/640	68/701

Class	Prod. Acc. (Percent)	User Acc. (Percent)	Prod. Acc. (Pixels)	User Acc. (Pixels)
-------	-------------------------	------------------------	------------------------	-----------------------

Trilhos T	47.92	68.49	150/313	150/219
Solo Nu T	93.48	99.63	1075/1150	1075/1079
Terreno Urban	95.99	91.82	1078/1123	1078/1174
Áreas Verdes	99.95	99.36	36353/36371	36353/36587
Água T	99.81	100.00	21224/21265	21224/21224
Vias Artifici	90.30	98.91	633/701	633/640

vii. Matriz de confusão árvore de decisão com dados auxiliares

Overall Accuracy = (51799/53897) 96.1074%

Kappa Coefficient = 0.9276

Ground Truth (Pixels)					
Class	Trilhos V	Solo Nu V	Áreas Verdes	Água V	Vias Artificiais
Unclassified	0	0	0	0	0
Trilhos	78	8	48	0	366
Solo Nu	129	1140	1336	25	93
Áreas Verdes	10	0	29103	0	1
Água	0	0	0	21240	0
Vias artificiais	62	2	18	0	238
Total	279	1150	30505	21265	698

Ground Truth (Pixels)	
Class	Total
Trilhos	500
Solo Nu	2723
Áreas Verdes	29114
Água	21240
Vias artificiais	320
Total	53897

Ground Truth (Percent)					
Class	Trilhos	Solo Nu	Áreas Verdes	Água	Vias Artificiais
Unclassified	0.00	0.00	0.00	0.00	0.00
Trilhos	27.96	0.70	0.16	0.00	52.44
Solo Nu	46.24	99.13	4.38	0.12	13.32
Áreas Verdes	3.58	0.00	95.40	0.00	0.14
Água	0.00	0.00	0.00	99.88	0.00
vias	22.22	0.17	0.06	0.00	34.10
Total	100.00	100.	100.00	100.00	100.00

Ground Truth (Percent)	
Class	Total
Trilhos	0.93
Solo Nu	5.05
Áreas Verdes	54.02
Água	39.41

Vias artificiais	0.59
Total	100.00

Class	Commission (Percent)	Omission (Percent)	Commission (Pixels)	Omission (Pixels)
Trilhos	84.40	72.04	422/500	201/279
Solo Nu	58.13	0.87	1583/2723	10/1150
Áreas Verdes	0.04	4.60	11/29114	1402/30505
Água	0.00	0.12	0/21240	25/21265
Vias artificiais	25.63	65.90	82/320	460/698

Class	Prod. Acc. (Percent)	User Acc. (Percent)	Prod. Acc. (Pixels)	User Acc. (Pixels)
Trilhos	27.96	15.60	78/279	78/500
Solo Nu	99.13	41.87	1140/1150	1140/2723
Áreas Verdes	95.40	99.96	29103/30505	29103/29114
Água	99.88	100.00	21240/21265	21240/21240
Vias Artificiais	34.10	74.38	238/698	238/320